Fourth order schemes for wave equation Maxwell's equations and linearized elastodynamic equations

Jukka Tuomela

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 1337

Programme 7
Calcul Scientifique,
Logiciels Numériques et Ingénierie Assistée
par Ordinateur

FOURTH ORDER SCHEMES FOR WAVE EQUATION MAXWELL'S EQUATIONS AND LINEARIZED ELASTODYNAMIC EQUATIONS

Jukka TUOMELA

Novembre 1990

RR-1337

Fourth Order Schemes for Wave Equation, Maxwell's Equations and Linearized Elastodynamique Equations

Schémas d'ordre quatre pour l'équation des ondes, les équations de Maxwell et les équations d'elastodynamique linéaire

Jukka Tuomela

Helsinki University of Technology
02150 Espoo
Finland

INRIA, Rocquencourt
BP 105
78153 Le Chesnay Cedex
France

October, 1990

Abstract

We analyse the effect of numerical dispersion for the wave equation, Maxwell's equations and linearized elastodynamic equations. We then construct some fourth order finite difference and finite element schemes for these equations using the modified equation approach and show that these are more efficient than ordinary second order schemes. Finally we present some results of the direct simulation of the two dimensional wave equation.

Resumé

Nous analysons la dispersion des schémas numériques pour l'équation des ondes, les équations de Maxwell et les équations d'elastodynamique linéaire et construisons des schémas de différences finies et d'éléments finis d'ordre quatre en espace et en temps pour ces mêmes équations. On montre que les schémas d'ordre quatre, bien que plus compliqués, sont plus efficaces que les schémas d'ordre deux. Ensuite on présente quelques résultats numériques pour l'équation des ondes en dimension deux.

1 Introduction

In the direct simulation of the wave propagation phenomena using finite elements or finite differences the major difficulty is the numerical dispersion, that is the signal propagates with the wrong speed and to the wrong direction [TR], [JO], [BCL]. The error in the direction of propagation is due to the fact that we replace the isotropic continuous problem by a discrete anisotropic problem (or if the original problem is anisotropic, the anisotropy of the discrete problem is different). Because with hyperbolic equations there is no damping of the error the quality of the solution reduces rather quickly as the time increases. Consequently constructing suitable high order schemes becomes interesting or even necessary (especially for three dimensional problems), the solutions being usually not too singular.

We will first consider the two dimensional wave equation and analyse the dispersion properties of some usual finite element and finite difference schemes. A disappointing fact is that if one wants an explicit scheme with P2 elements (or any other ordinary PN elements) this reduces the order of the method to two even in space variables. Next we give necessary and sufficient conditions which guarantee that the artificial numerical anisotropy is small. With the aid of these low anisotropy schemes, it is then possible to construct methods which are of fourth order both in space and in time using the modified equation approach (see [CJ], [CO], [SB]). We suppose that the signal speed is locally constant, but the generalization to smoothly varying speed may be possible. Anyway the attractive feature of these schemes is their easy implementation and their rapidity. We compare the different schemes and show that the fourth order schemes are superior to the second order ones.

Extending the above ideas we construct a fourth order finite difference scheme for three dimensional wave equation. With finite elements we have then the fundamental geometrical problem of filling the space with tetrahedra. Now to have a fourth order scheme using the same construction as before these tilings have to be quite symmetric, and it seems that it is impossible to have 'enough' symmetry. Curiously, it appears still to be an open problem to classify or describe all the different tilings with (more or less) regular tetrahedra, [GO], [GS]. Anyway we calculate the dispersion relation for the most regular tiling and show that it is not symmetric enough to give a fourth order scheme. These considerations suggest that except P1 elements in two dimensions, the standard finite elements are not really suitable for the wave propagation calculations.

Then we pass to the Maxwell's equations and linearized elastodynamic equations. We construct a family of fourth order methods for both of these equations. The qualitative behavior of the solutions being essentially the same as with the simple wave equation we have exactly the same reasons to use fourth order methods as before. There are, however, some additional features: the new equations being vector valued the discretisation changes not only the magnitude of this vector but its direction as well (polarization error), very much in the same way as not only the speed but

also the direction of propagation changes. For our methods this polarization error is small.

Elastodynamic equations are still more complicated: there are two kinds of waves, pressure wave and shear wave, and their speed is different. However, using our method the pressure wave behaves exactly like a solution to the three dimensional wave equation so it is enough to consider shear waves. Now the speed of the shear wave appears as a double eigenvalue of some operator and a nice feature of our scheme is that the discretised operator also has a double eigenvalue (and not two eigenvalues close to each other). It is seen that the error grows quite quickly as a function of the ratio speed of the pressure wave / speed of the shear wave. This seems to be difficult to avoid with any scheme.

Finally we present some numerical simulations in two dimensions for the wave equation. Using the known analytical solution we then compare the accuracy of different methods and the time needed to compute the solution. The results show that it pays to use fourth order methods even if the solution is not sufficiently smooth for the error to be $O(h^4)$.

2 Two Dimensional Wave Equation

2.1 Numerical Dispersion

Let us start by considering the ordinary two dimensional wave equation

$$u_{tt} - c^2 \Delta u = 0 \tag{2.1}$$

Here c is the speed of the propagation, which for the simplicity of the notation we take to be one. The equation (2.1) has then the plane wave solutions

$$u = e^{i(k \cdot r - \omega t)} \tag{2.2}$$

with the dispersion relation

$$\omega^2 = |k|^2 \tag{2.3}$$

where ω is the angular frequency (or simply frequency), k is the wave vector, $|k|$ the wave number and r is the vector $(\ x \quad y\ )$ (or $(\ x \quad y \quad z\ )$ in three dimensional space). The wave vector points to the direction of the propagation or in other words it is perpendicular to the wavefront. If we discretise (2.1) with finite differences using the usual five point scheme in space and three point scheme in time we get the following difference equation

$$\frac{u_{i,j}^{n+1} - 2u_{i,j}^n + u_{i,j}^{n-1}}{\delta t^2} + \frac{1}{h^2}(4u_{i,j}^n - u_{i+1,j}^n - u_{i-1,j}^n - u_{i,j+1}^n - u_{i,j-1}^n) = 0 \tag{2.4}$$

Here δt is the time step and h is the mesh parameter. There are plane wave solutions provided that the following dispersion relation is satisfied.

$$\frac{4}{\delta t^2}\sin^2(\omega\delta t/2) = \frac{4}{h^2}(\sin^2(k_1 h/2) + \sin^2(k_2 h/2)) \tag{2.5}$$

In order to have a real ω we impose the following stability condition

$$\alpha^2(\sin^2(a_1/2) + \sin^2(a_2/2)) \leq 1$$

where we denote $k_i h$ by a_i and $\delta t/h$ by α. Let us note that if $c \neq 1$ then all our results remain valid if one replaces δt by $c\,\delta t$ and consequently then $\alpha = c\,\delta t/h$. This leads to

$$\alpha \leq 1/\sqrt{2} \simeq 0.71$$

Then using the Taylor's series we get

$$\omega^2 = |k|^2(1 + (\delta t^2 - \frac{k_1^4 + k_2^4}{|k|^4}h^2)|k|^2/12 + O(h^4 + \delta t^4))$$

Remebering that the phase speed is defined by $v_\varphi = \frac{\omega}{|k|^2}k$ we notice that the numerical phase speed is bigger along the diagonals ($k_1 = \pm k_2$) than along the mesh ($k_1 = 0$ or $k_2 = 0$). In the original equation the speed is obviously independent of the direction; the numerical scheme has introduced anisotropy into the system. In fact more important concept in the wave propagation is the group speed defined by $v_g = \nabla\omega$, where the gradient is taken with respect to k. This is because the energy in the system propagates (approximately) with the group speed, [BCL], [TR]. Note that for the continuous problem (2.1) the phase speed and the group speed is the same thing (take the gradient of (2.3)); on the other hand for the discrete problem their lenght and direction can be (and usually are) different. The error in the group speed is typically greater than in the phase speed. We illustrate this in the case of one space dimension. Suppose that there is no error in the time discretisation, so that the dispersion relation of any second order scheme has the following form.

$$\omega^2 = k^2(1 - 2ak^2h^2 + O(h^4))$$

Then simple calculations show that

$$\begin{aligned} v_\varphi = \omega/k = &\quad 1 - ak^2h^2 + O(h^4) \\ v_g = d\omega/dk = &\quad 1 - 3ak^2h^2 + O(h^4) \end{aligned}$$

So the error in the group speed is about three times as big as in the phase speed (note that this does not depend on the dimension of the space). Similarly for the fourth order schemes we get the factor five for the errors. However the order of the error is evidently the same.

2.2 Mass Lumping

Let us then turn to the finite elements. We will only consider different triangular elements, so the grid is composed entirely of triangles. Denoting by M the mass matrix and by R the stiffness matrix, we can conveniently write the time discretisation as follows.

$$M\frac{u^{n+1}-2u^n+u^{n-1}}{\delta t^2}+R(\theta u^{n+1}+(1-2\theta)u^n+\theta u^{n-1})=0$$

Here $\theta \in [0,1/2]$ is a free parameter and u^n is a vector whose components give the values of the degrees of freedom at time $t=n\delta t$. If $\theta \in [1/4,1/2]$ then the above scheme is stable for any δt. However the scheme is then implicit, so we must solve a linear system at every time step. This takes a lot of time and it is not really necessary: the stability condition for the hyperbolic equations is always reasonable (this is due to the fact that the signal speed is finite). So we would like to take $\theta = 0$; but this does not still in general yield an explicit method because usually the mass matrix is not diagonal. Consequently we want to replace the mass matrix by a diagonal matrix (mass-lumping) without reducing the order of accuracy .

There are two ways of doing this (for details, see [ZI] and [DL1]). First we can use a suitable numerical integration method to evaluate the mass matrix. On the other hand to preserve the accuracy the quadrature must satisfy some conditions. Consider the general elliptic problem

$$-\nabla\cdot(C\nabla u)=f$$

Then the quadrature has to be accurate enough to evaluate exactly the stiffness matrix, when the tensor C in the above equation is constant. In our case this simply means that when the speed of propagation is constant, the stiffness matrix is evaluated exactly. Note that taking the quadrature points to be the lagrangian interpolation points diagonalizes the mass matrix of any PN element, but it satisfies the above accuracy rule only for P1 elements (the quadrature evaluates exactly the first degree polynomials). We remark that the same effect is obtained if one takes the sum of every row of the mass matrix and puts these sums on the diagonal and zeros everywhere else. Denoting by $\{\varphi_i\}_{i=1}^N$ the family of basis functions and by d_j the diagonal element in the j'th row after mass-lumping we can write

$$d_j=\sum_{i=1}^{N}\int\varphi_i\varphi_j=\int\varphi_j$$

One can express this by saying that d_j represents the 'mass' of φ_j.

The second possibility is simply to use different basis functions for mass matrix than for the stiffness matrix. This is of course more or less arbitrary but at least in some cases it can be made to work. Obviously we want that the new basis functions are

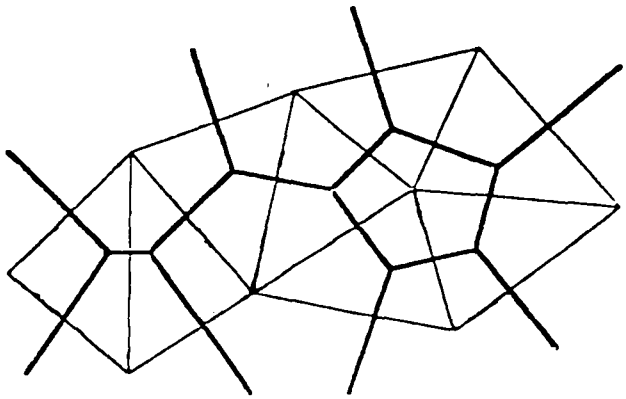


Figure 2.1: The underlying triangulation and its Voronoi diagram

as simple as possible, for instance piecewise constants. So we associate to the triangulation its Voronoi diagram, see figure 2.1, and define the basis function assosiated to any point to be one inside its Voronoi polygon and zero outside. If we choose equilateral triangles, then the support of the new basis function will be a regular hexagon surrounding the corresponding point. In case of PN elements we associate to the original mesh a finer mesh (PN iso P1) whose vertices are the lagrangian interpolation points of the original mesh and then construct the Voronoi diagram with respect to this new mesh. With this strategy the diagonal element of every point is the same (as long as the original triangulation is regular), which is rather natural. Notice that if we use this trick with the (conforming) P1 element, we get identical results as obtained above by the numerical integration method.

2.3 P1 element

Let us then consider ordinary conforming P1 element. First note that if we choose the triangulation as in figure 2.2 and use mass lumping we get exactly the same dispersion relation as in (2.5). To proceed we need some notations. Let us first define the mesh vector $\bar{h}$. With finite elements the orientation of the grid is independent of the coordinate system and we must be able to represent different orientations in some way. So we take one edge of the grid and associate to it the vector $\bar{h}$ and then express all other vectors needed with aid of $\bar{h}$. Then define $h = |\bar{h}|$. Finally let A be

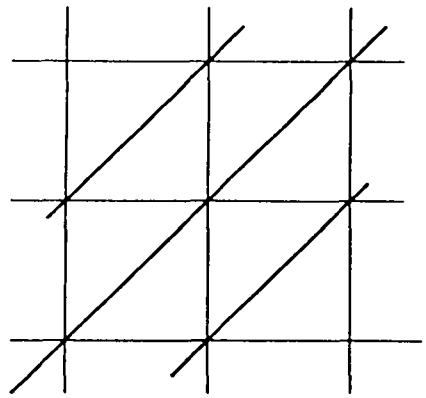


Figure 2.2: The rectangular triangulation.

a rotation by 60 degrees, that is

$$A = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix}$$

As noted before the problem with the method (2.4) is that it is anisotropic. We try to improve the situation by taking a mesh which is more symmetric: we tile the plane with equilateral triangles. We use the mass lumping and try to find the resulting dispersion relation. Because the degrees of freedom and the lagrangian interpolation points coincide, we can interprete the resulting scheme as a certain finite difference scheme and use the same analysis by plane waves as above.

We use the same three point time discretisation as before. Then after some straightforward but rather long calculations (these are most conveniently done using barycentric coordinates, see [RT]), we get the following difference scheme.

$$\frac{u_0^{n+1} - 2u_0^n + u_0^{n-1}}{\delta t^2} + \frac{2}{3h^2}(6u_0^n - \sum_{l=1}^{6} u_l^n) = 0 \tag{2.6}$$

The points u_l are the six neighbors of the node u_0. For later purposes we define the difference operator L.

$$L\,u = \frac{2}{3}(6u - \sum_{l=1}^{6} u_l) \tag{2.7}$$

Substituting the plane wave (2.2) into (2.6) we find the following dispersion relation.

$$\frac{4}{\delta t^2}\sin^2(\omega\delta t/2) = \frac{8}{3h^2}(\sin^2(k\cdot\bar{h}/2) + \sin^2(k\cdot A\bar{h}/2) + \sin^2(k\cdot A^2\bar{h}/2))$$

Denote the right hand side multiplied by h^2 in the above equation by f_L. Then we can write the stability condition as

$$\alpha^2 f_L \leq 4 \tag{2.8}$$

Obviously f_L depends only on k and the angle between k and $\bar{h}$ so without the loss of generality we can choose $\bar{h} = (\, h \quad 0\,)$. Then using the elementary differential calculus we maximize f_L with respect to k. This gives

$$\max_{k \in \mathbb{R}^2} f_L(k) = 6$$

Substituting this into (2.8) we finally get

$$\alpha \leq \sqrt{\frac{2}{3}} \simeq 0.82 \tag{2.9}$$

Using the Taylor's series as usual we find

$$\omega^2 = |k|^2(1 + (4\delta t^2 - 3h^2)|k|^2/48 + O(h^4 + \delta t^4)) \tag{2.10}$$

Now the second order term is isotropic so we have a fourth order scheme if we choose

$$\alpha = \delta t/h = \sqrt{3}/2 \simeq 0.87$$

Comparing to (2.9) we notice that this unfortunately does not satisfy the stability condition.

2.4 Non Conforming P1 Element

Next let us consider the non conforming P1 element. It has the curious property that the mass matrix is exactly diagonal (this is not true in three dimensions), so we don't have to worry about mass lumping. There is however another difficulty: the points in different edges are not equivalent, see figure 2.3. More precisely we say that points a and b are equivalent if there exists a translation of the space which maps a to b and the neighbors of a to the neighbors of b. Consequently one has to solve a three dimensional eigenvalue problem to get the dispersion relation and the stability condition. Moreover we get not only the the solution which approximates the dispersion relation (2.3), but also two other solutions, which correspond to (highly oscillating) parasitic waves. In general if we have n inequivalent points we have only one 'good' solution and $n - 1$ parasitic solutions.

Let us again choose the same triangulation with equilateral triangles as before. We change the meaning of h slightly: to be able to easily compare different methods, h will be here half of the length of the edge (and a similar change for $\bar{h}$). Then for all

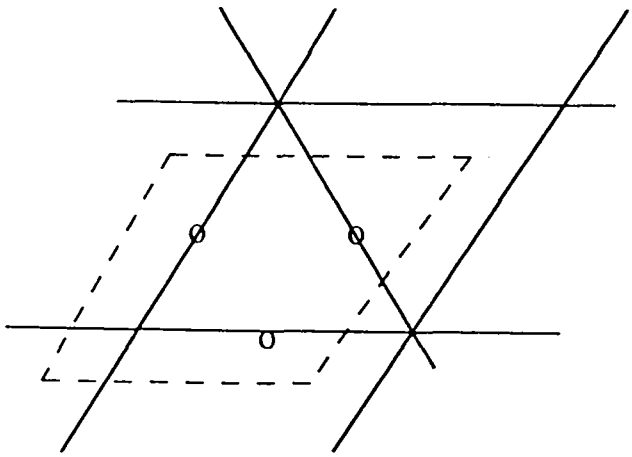


Figure 2.3: Three points in the 'unit rectangle' are inequivalent.

schemes considered we can say that the same h means: the shortest distance between the Lagrangian interpolation points (or about the same number of the degrees of freedom in the given area).

Proceeding as before we get the following dispersion relation (or eigenvalue problem)

$$\frac{4}{\delta t^2}\sin^2(\omega\delta t/2)x = \frac{1}{h^2}B_{nc}x \tag{2.11}$$

where B_{nc} is given by

$$B_{nc} = 2\begin{pmatrix} 2 & -\cos(k\cdot A^2\bar{h}) & -\cos(k\cdot A\bar{h}) \\ -\cos(k\cdot A^2\bar{h}) & 2 & -cos(k\cdot\bar{h}) \\ -\cos(k\cdot A\bar{h}) & -cos(k\cdot\bar{h}) & 2 \end{pmatrix}$$

When $k = 0$, we see that

$$B^0_{nc} = \begin{pmatrix} 4 & -2 & -2 \\ -2 & 4 & -2 \\ -2 & -2 & 4 \end{pmatrix}$$

The matrix B^0_{nc} has the eigenvalues 6, 6 and 0. So we get immediately the lower bound for the biggest eigenvalue λ_b of B_{nc}/h^2 (remember that we really take the maximum over all values of k).

$$\lambda_b \geq \frac{6}{h^2}$$

On the other hand the classical Geršgorin's theorem (see [HJ]) gives the upper bound

$$\lambda_b \leq \frac{9}{h^2}$$

So denoting by α^s the stability limit we have the following inequalities.

$$0.67 \simeq 2/3 \leq \alpha^s \leq \sqrt{\frac{2}{3}} \simeq 0.82$$

The eigenvalue λ_k, which approximates $|k|^2$ can again be found using Taylor's series, which leads to the following dispersion relation.

$$\omega^2 = |k|^2(1 + (\delta t^2 - h^2)|k|^2/12 + O(h^4 + \delta t^4))$$

Again we note that the second order term is isotropic and the 'optimal' time step $\delta t = h$ is unstable.

It can be verified that the eigenvector corresponding to λ_k tends to $(1 \quad 1 \quad 1)$ as $h \to 0$, which is natural and desirable. On the other hand the eigenspace corresponding to other eigenvalues tends to the subspace spanned by $(1 \quad -1 \quad 0)$ and $(1 \quad 0 \quad -1)$. So the sum of the components is always zero. This can be interpreted by saying that the parasitic waves tend weakly to zero as $h \to 0$.

2.5 P2 Element

Next we treat the P2 element. We recall that the grid is the same as before and h is half of the length of the edge. Here we have four inequivalent points: the vertex and the points on the three edges (see figure 2.4). The problem with P2 elements is that we cannot as easily diagonalize the mass matrix as before: if we use a numerical quadrature which satisfies the rule formulated above (the quadrature points are in the middle of the edges, see [DL1]), we get a diagonal matrix, but in addition the diagonal elements corresponding to the vertex points become zero (the basis function corresponding to these points has 'zero mass', that is its integral vanishes). Note that this does not depend on the geometrical properties of the mesh. Of course these zeros are inacceptable because we have to divide by the diagonal. To overcome this difficulty we then use the second method of mass lumping mentioned above.

Otherwise we proceed in the same way as in the precceding case. The dispersion relation is then of the same form as (2.11), where B_{nc} is replaced by

$$B_{P2} = \frac{16}{9}\begin{pmatrix} 3(1 - f_L/16) & -\cos(k \cdot \bar{h}) & -\cos(k \cdot A^2\bar{h}) & -\cos(k \cdot A\bar{h}) \\ -\cos(k \cdot \bar{h}) & 3 & -\cos(k \cdot A\bar{h}) & -\cos(k \cdot A^2\bar{h}) \\ -\cos(k \cdot A^2\bar{h}) & -\cos(k \cdot A\bar{h}) & 3 & -\cos(k \cdot \bar{h}) \\ -\cos(k \cdot A\bar{h}) & -\cos(k \cdot A^2\bar{h}) & -\cos(k \cdot \bar{h}) & 3 \end{pmatrix}$$

Here f_L is the same as in the P1 case except we have to replace h by $2h$. Taking $k = 0$ we find

$$B^0_{P2} = \frac{16}{9}\begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix}$$

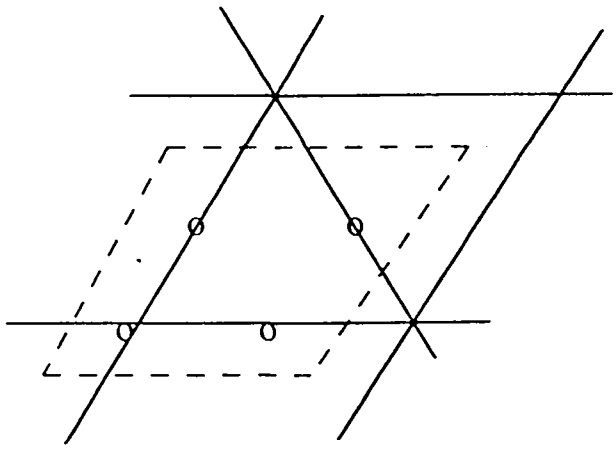


Figure 2.4: The four points in the 'unit rectangle' are inequivalent.

Three of the eigenvalues of B^0_{P2} are equal to 64/9 and the fourth is zero. Again this and the Geršgorin's theorem gives us the following bounds for the biggest eigenvalue λ_b of B_{P2}/h^2.

$$\frac{64}{9h^2} \leq \lambda_b \leq \frac{32}{3h^2}$$

Consequently we get the following stability condition, denoting by α^s the stability limit as before.

$$0.61 \simeq \sqrt{\frac{3}{8}} \leq \alpha^s \leq \frac{3}{4} = 0.75 \tag{2.12}$$

Next expanding the small eigenvalue of B_{P2} we find

$$\omega^2 = |k|^2(1 + (16\delta t^2 - 9h^2)|k|^2/192 + O(h^4 + \delta t^4))$$

The optimal time step is then seen to be $\delta t = 3h/4$, which is exactly at the upper bound of the stability limit. In practise this limit cannot quite be attained. The eigenvectors behave as in the non conforming case; the eigenvector corresponding to the desired solution tends to (1 1 1 1) and if a vector belongs to the eigenspace of the rest, then the sum of its components tends to zero.

We remark that the accuracy is of the second order even in space variables although we are using $P2$ elements; mass lumping has not preserved the accuracy. This being the case we could also diagonalize the mass matrix by 'quadrature points are interpolation points' method. To analyse this situation we simply multiply B_{P2} by a diagonal matrix whose diagonal elements are (1/2 3/2 3/2 3/2). Note that the

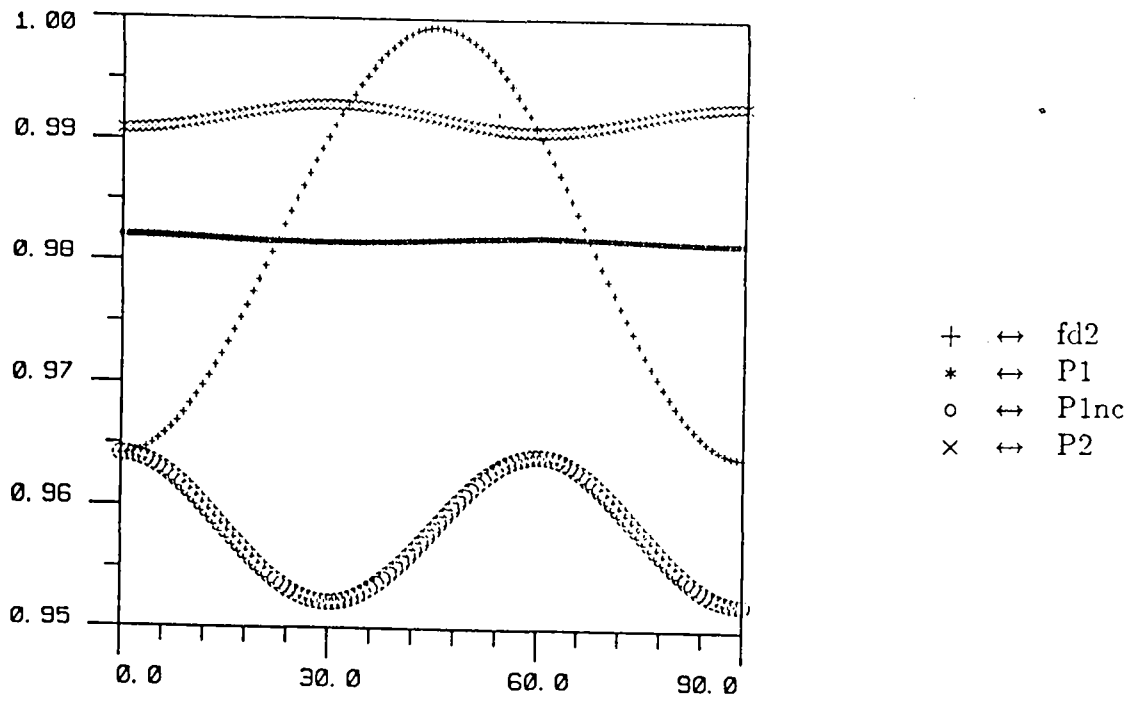


Figure 2.5: Comparison of the second order schemes

resulting matrix is not symmetric anymore. Doing the same calculations as above we find the following stability bounds and Taylor's expansion.

$$0.5 = 1/2 \leq \alpha^* \leq \sqrt{\frac{3}{8}} \simeq 0.61$$

$$\omega^2 = |k|^2(1 + (4\delta t^2 - 9h^2)|k|^2/48 + O(h^4 + \delta t^4))$$

The error coefficient in space discretisation is here four times bigger than in the previous case, so we can conclude that the first way to diagonalize the mass matrix is better. Also the stability condition is less strict in that case.

To get an idea of the accuarcy of the above basic schemes let us plot the phase speed for some parameter values. We note in passing that to minimise the work one would obviosly like to choose α as big as possible and fortunately this is not incompatible with accuracy requirements: looking at the dispersion relations for the different methods it is seen that typically the accuracy improves as α grows; the error terms associated to space and time discretisations are of different sign.

In figure 2.5 we have then the phase speed as a function of the direction of the wave vector ($|k| = 1$ here and in all the pictures that follow: it is just a matter of scaling). For every scheme we have chosen $\alpha = 0.7$ for simplicity. The phase speed should be

one because we chose $c = 1$ in (2.1). We recall that the wavelength is $2\pi/|k|$ and number of points per wavelength $N = 2\pi/h|k|$ (N is not necessarily an integer). In figure 2.5 we have taken $N = 6$. We also note that according to the famous theorem of Shannon to represent signal correctly we must at least have $N > 2$. Because it is the maximum of the error which is important we see that the P2 elements are the most accurate.

3 Direction of Propagation

As we remarked above the error in the absolute value of group speed is about five times bigger than in phase speed for the fourth order scheme so it is not necessary to make separate plots for the group speed. However, there is also the direction of the propagation that has to be taken into consideration. We recall that for the original solution the group speed and the wave vector are parallel, so a natural way to mesure the error is then to calculate

$$\beta = \arccos\left(\frac{k \cdot \nabla\omega}{|k|\,|\nabla\omega|}\right)$$

Let us first note that β does not depend on the time discretisation; in fact for the method (2.6) (and similarly for other schemes) we have

$$\frac{k \cdot \nabla\omega}{|k|\,|\nabla\omega|} = \frac{k \cdot \nabla f_L}{|k|\,|\nabla f_L|} \tag{3.1}$$

Next, it is straightforward to verify that if the scheme is of the order m then

$$\begin{aligned} \frac{k \cdot \nabla\omega}{|k|\,|\nabla\omega|} &= 1 - bh^{2m} + O(h^{2m+1}) \\ \beta &= \sqrt{2b}h^m + O(h^{m+1/2}) \end{aligned}$$

where b is some positive constant. So β is then normally of the same order as the order of the scheme. However, in some cases the order of β can be higher than the order of the method. To state the conditions under which this is possible let us introduce some notation. First we write any dispersion relation generically as

$$\frac{4}{\delta t^2}\sin^2(\omega\delta t/2) = g/h^2$$

In view of (3.1) it is then enough to consider g. To avoid repetitions we state the conditions directly in three dimensions. Now there are some conditions which any

reasonable g has to satisfy: for any consistent scheme we have

$$\begin{aligned} &(C1) \quad g(0,0,0) = 0 \\ &(C2) \quad \partial g/\partial k_i = 0 \\ &(C3) \quad \partial^2 g/\partial k_i^2 = 2 \\ &(C4) \quad \partial^2 g/\partial k_i \partial k_j = 0 \end{aligned}$$

In (C2) - (C4) the partial derivatives are evaluated at the origin and the index i ranges from one to two or three, according to the dimension of the space. In (C4) obviously $i \neq j$. In (C3) the constant is 2, because the signal speed is one, otherwise we would have $2\,c^2$. Then we have a condition which is not indispensable, but which is extremely natural in the context of wave propagation phenomena, namely

$$(C5) \quad g(k_1, k_2, k_3) = g(|k_1|, |k_2|, |k_3|)$$

That is g is an even function of any of its arguments. It is easily seen that (C5) implies (C2) and (C4), and more generally it implies that any partial derivative which contains an odd order derivative is automatically zero at the origin. Now if we have an irregular finite element mesh, then evidently (C5) is not verified, but these kind of meshes are not really interesting because they just introduce parasitic waves into the solution without improving accuracy. Now we can formulate the following

Proposition 1 *Suppose that g satisfies conditions (C1) - (C5); then $\beta = O(h^4)$ if and only if for any choice of indices i, j and l ($j \neq l$)*

$$\frac{\partial^4 g}{\partial k_i^4} = 3\frac{\partial^4 g}{\partial k_j^2 \partial k_l^2} \tag{3.2}$$

where the derivatives are evaluated at the origin.

The interpretation of the above condition is that the second order error term does not depend on the direction of the wave vector k. The proof is of course a tedious verification, where we first suppose that (3.2) is not satisfied. Then it is rather easy to find such k, that $\beta = O(h^2)$ in that direction. Next supposing (3.2) to be true, we calculate that

$$\frac{k \cdot \nabla g}{|k|\,|\nabla g|} = 1 + O(h^8)$$

which gives the desired result. Now for the later purposes we formulate a similar, but slightly more general result. Consider a vector valued function v whose components satisfy the following conditions (we write the conditions for v_1, others verify similar conditions which are obtained by permutation of indices).

$$(D1) \quad v_1(k_1, k_2, k_3) = v_1(k_1, |k_2|, |k_3|)$$

$$(D2) \qquad v_1(k_1, k_2, k_3) = -v_1(-k_1, k_2, k_3)$$
$$(D3) \qquad \partial v_1/\partial k_1 = 1$$
$$(D4) \qquad \frac{\partial^3 v_1}{\partial k_1^3} = 3\frac{\partial^3 v_1}{\partial k_1 \partial k_i^2}$$

Evidently in (D4) $i \neq 1$ and the derivatives are evaluated at the origin. We then have the following result.

Proposition 2 *Suppose that v satisfies (D1) - (D3); then v satisfies (D4) if and only if*

$$\frac{k \cdot v}{|k|\,|v|} = 1 + O(h^8)$$

The proof is similar as in the previous case. The connection between the two propositions is that essentially $\partial g/\partial k_i$ is like v_i, but the proposition 2 is needed in another context later.

Next we give a sufficient condition for the property (3.2) in the two dimensional case.

Proposition 3 *Suppose that in addition to conditions (C1) - (C5) g is invariant by rotation of some angle $\neq n\,\pi/2$; then it satisfies* (3.2).

The proof is a straightforward verification. Note that f_L is invariant by rotation of $\pi/3$. Finally let us mention the following simple result.

Proposition 4 *In three dimensions g satisfies (3.2) if and only if it satisfies the two dimensional conditions for* $k = (\,k_1 \quad k_2 \quad 0\,)$, $k = (\,k_1 \quad 0 \quad k_3\,)$ *and* $k = (\,0 \quad k_2 \quad k_3\,)$.

Next we show the errors in the direction of the group speed for the usual five-point difference scheme and for the P1 elements with equilateral triangles. In figure 3.1 (where we have taken $N = 5$) the superiority of the P1 scheme is clear. The other values of N give similar results.

4 Compensation

4.1 Modified Equation Approach

The trouble with the above 'optimal' time steps (apart from the fact that they tend to give unstable schemes) is that they are not flexible; of course we would like to have a fourth order scheme for any sufficiently small time step. However, there is a fundamental difficulty with the time discretisation: if we use a centered fourth order approximation for the second time derivative, then the scheme is unconditionally unstable. On the other hand the centered approximation is necessary if we want to

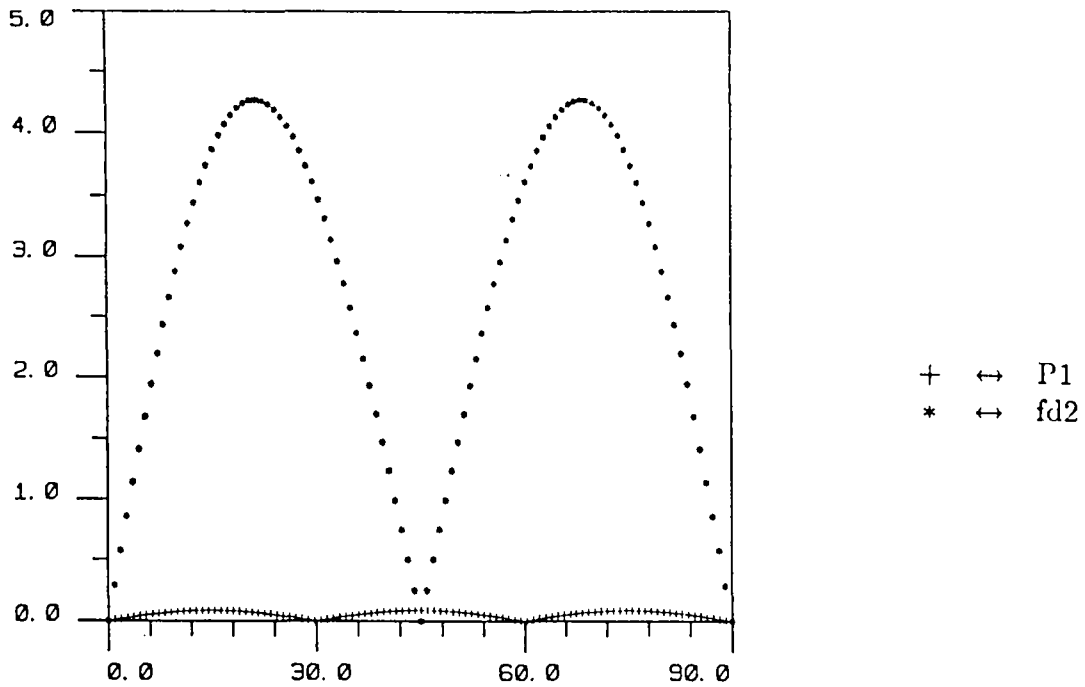


Figure 3.1: Error in the direction of propagation; $N = 5$

avoid introducing artificial dissipativity. There is however a way out. Consider the usual three point scheme.

$$\frac{u^{n+1} - 2u^n + u^{n-1}}{\delta t^2} = u_{tt} + \frac{\delta t^2}{12} u_{tttt} + O(\delta t^4)$$

Now using the equation (2.1) this becomes

$$\frac{u^{n+1} - 2u^n + u^{n-1}}{\delta t^2} = u_{tt} + \frac{\delta t^2}{12} \Delta^2 u + O(\delta t^4)$$

So instead of the original equation (2.1) we might solve the following modified equation (see [CJ], [CO], [SB])

$$u_{tt} - \Delta u - \frac{\delta t^2}{12} \Delta^2 u = 0 \tag{4.1}$$

Now if the discretisation of the laplacian is of the fourth order and the bilaplacian of the second order (note the factor $\delta t^2/12$), then the scheme is of the fourth order both in space and time. It is interesting to note that the equation (4.1) does not yield a well posed problem because of the non positivity of the last term. However, it is still possible to construct stable numerical schemes (intuitively this is rather natural: because δt is 'small', multiplying by it 'cancels' one derivative, so that numerically

$\delta t^2\Delta^2$ behaves like a second order differential operator). But then by the same argument one could also compensate the error made when discretising the laplacian, provided that the second order error term is isotropic. So let us again consider the P1 elements with equilateral triangles. Then using the operator defined in (2.7) we have

$$\frac{1}{h^2}L\,u = -\Delta u - \frac{h^2}{16}\Delta^2 u + O(h^4)$$

Proving this is just the same calculation that lead to (2.10). But for the idea to be useful we have to approximate the bilaplacian in some way. However, this is simple because

$$\frac{1}{h^4}L^2 u = \Delta^2 u + O(h^2)$$

and as noted above the second degree accuracy is enough. This leads us to the equation

$$u_{tt} - \Delta u + \frac{3h^2 - 4\delta t^2}{48}\Delta^2 u = 0$$

We remark that for small δt the coefficient of the last term is positive and consequently we have a well posed problem in an appropriate Sobolev space (although this is not really important in practise, as noted above). The numerical scheme is then simply

$$\frac{u^{n+1} - 2u^n + u^{n-1}}{\delta t^2} + \frac{1}{h^2}\left(Lu^n + \frac{3h^2 - 4\delta t^2}{48h^2}L^2 u^n\right) = 0 \tag{4.2}$$

So all that is needed is one extra matrix vector multiplication at each time step. Obviously the numerical scheme is stable for small enough δt. The exact stability condition is also easy to calculate. Let us first write the dispersion relation for (4.2). Using the same f_L as in (2.8) we get

$$\frac{4}{\delta t^2}\sin^2(\omega\delta t/2) = \frac{1}{h^2}\left(f_L + \frac{9h^2 - 12\delta t^2}{144h^2}f_L^2\right) \tag{4.3}$$

The stability condition is then

$$\alpha^2\left(f_L + \frac{9 - 12\alpha^2}{144}f_L^2\right) \leq 4$$

Next we note that the left hand side attains its maximum at the same value of k as f_L (f_L is always positive and the coefficient of the f_L^2 is positive for the values of α which interest us). We recall that the maximum value of f_L was 6. Then making the substitutions we get the following inequality

$$12\alpha^4 - 33\alpha^2 + 16 \geq 0$$

Solving the above inequality yields

$$\alpha \leq \sqrt{\frac{33-\sqrt{321}}{24}} \simeq 0.79$$

Note that this is not really much stricter than (2.9). In fact the compensation can sometimes even improve stability as we will see later. This is consistent with the observation above that the added term is not really a fourth order operator: if it were, the stability condition would be much stricter (in general one would have $\delta t = O(h^2)$).

Of course exactly the same idea can be applied to P2 and the non conforming P1 elements. Proceeding as in P1 case we have instead of (4.3) the following eigenvalue problems.

$$\begin{aligned}
\frac{4}{\delta t^2}\sin^2(\omega\delta t/2)x &= \frac{1}{h^2}\left(B_{nc}+\frac{h^2-\delta t^2}{12h^2}B_{nc}^2\right)x \\
\frac{4}{\delta t^2}\sin^2(\omega\delta t/2)x &= \frac{1}{h^2}\left(B_{P2}+\frac{9h^2-16\delta t^2}{192h^2}B_{P2}^2\right)x
\end{aligned}$$

As before we calculate only the lower and upper bounds of the stability conditions, which gives

$$\begin{aligned}
0.54 \simeq \sqrt{\frac{21-\sqrt{249}}{18}} \leq \quad \alpha_{nc}^s \leq \quad \sqrt{\frac{9-\sqrt{33}}{6}} \simeq 0.74 \\
0.55 \simeq \sqrt{\frac{27-3\sqrt{33}}{32}} \leq \quad \alpha_{P2}^s \leq \quad 3/4 = 0.75
\end{aligned}$$

We could also have used the other P2 mass lumping which was mentioned in the previous section. However, it turned out to be much less accurate than the 'normal' P2 method, so we will not consider it.

4.2 Comparing the Accuracy of Different Methods

We will next present some curves of the dispersion relations of the different schemes to be able to compare their accuracy. First in figure 4.1 (respectively 4.2 and 4.3) the phase speed is shown as a function of $2/N$ (N is the number of points per wavelength) and the different curves correspond to different values of α for P1 (resp. P2 and non conforming P1) elements. In each of the figures the direction of k was chosen as $\varphi = 20$. We see that accuracy increases monotonically with respect to α (α goes from zero to the stability limit) which is nice because it saves computing time. Consequently in each case it is best to choose α exactly at the stability limit.

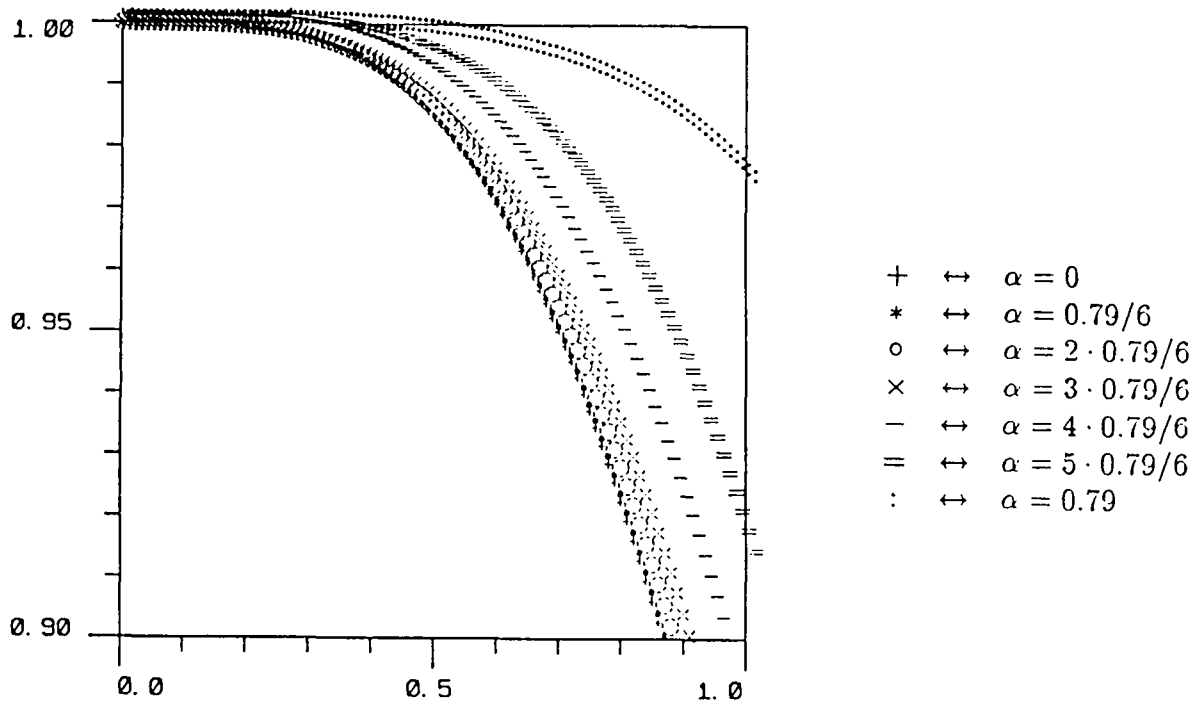


Figure 4.1: The phase speed for P1 elements

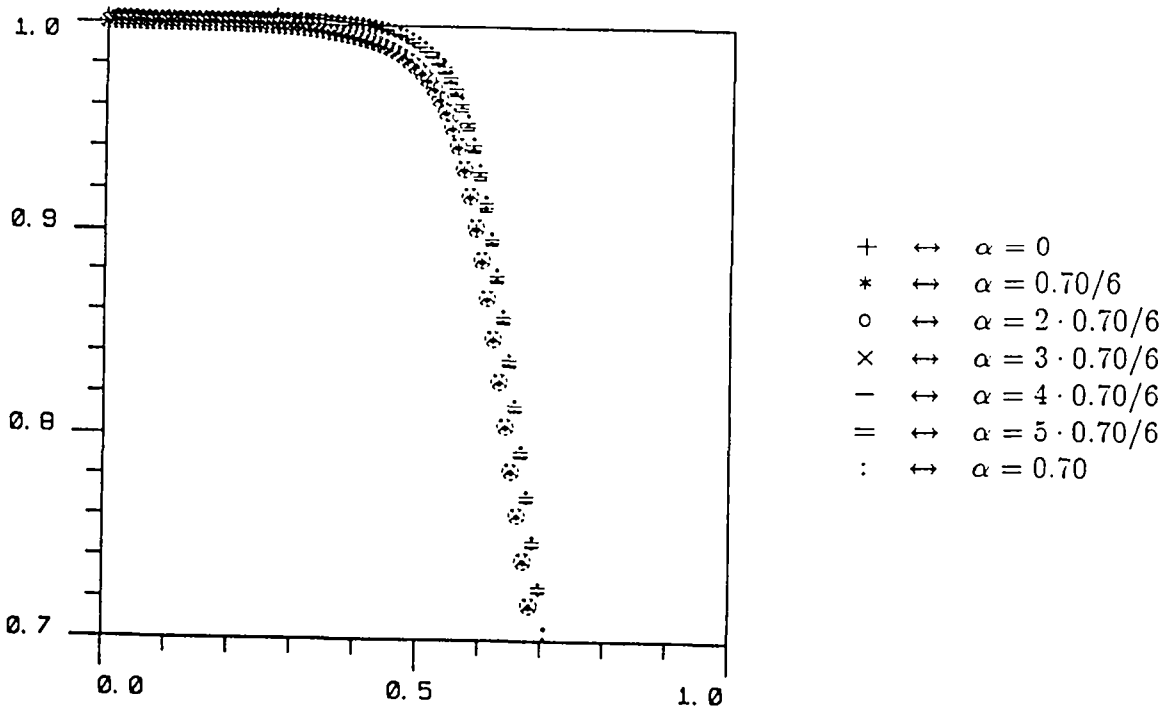


Figure 4.2: The phase speed for P2 elements

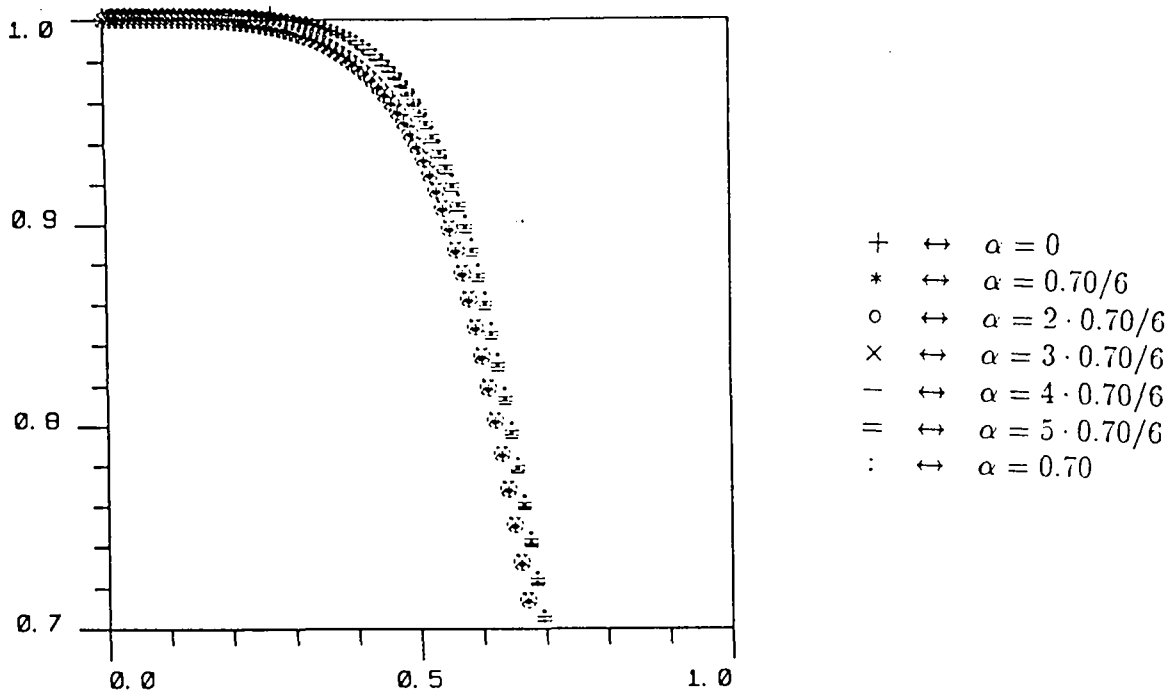


Figure 4.3: The phase speed for non conforming P1 elements

In figure 4.4 we then compare the different elements, where for each method α was chosen to be 0.7 and $\varphi = 20$ as before. Obviously if $N < 4$ P1 elements are the most accurate. Let us note that we have not calculated the exact stability limit for P2 and non conforming P1 elements and instead we have used the 'reasonable' value 0.7 for both of them. The conclusions would not have been changed if we had chosen for instance 0.75 for them.

Next let us plot the equivalent of figure 2.5 for different methods choosing again the same $\alpha = 0.7$ and the same mesh parameter in each case. In figure 4.5 we take $N = 5$ and in figure 4.6 $N = 10$. From the above figures we see that the non conforming P1 scheme is clearly the worst of the above fourth order methods. The P1 and P2 schemes give about the same precision, P1 being slightly better. However, this comparison does not take into account the effect of the parasitic waves, so in practice we can expect that P1 scheme is quite a lot better than P2 scheme.

Finally from computational point of view P2 elements are more expensive, because the stiffness matrix is fuller than in P1 case. This, combined with the above results, shows that it does not pay to use P2 elements with the hyperbolic equations. Intuitively this is perhaps not so surprising: first the use of mass lumping is 'unnatural' with P2 elements or in other words the basis functions are not well adapted to mass lumping. Also the fact that all points are not equivalent means 'physically' that the continuous homogeneous material has been replaced by discrete inhomogeneous material; this creates the parasitic waves.

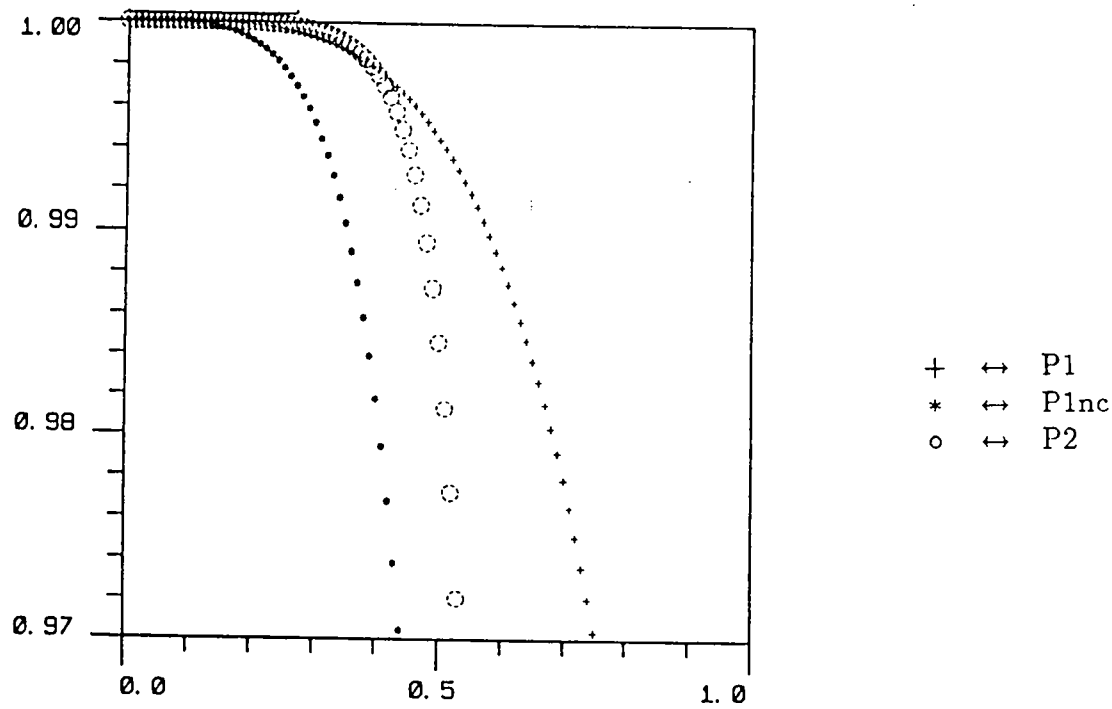


Figure 4.4: Comparison of different methods; $\varphi = 20$ and $\alpha = 0.7$

4.3 Comparing the Work

Of course fourth order schemes are more complicated and expensive than second order schemes, so to be able to compare them one should consider also the amount of work which has to be done. We will simply assume that the amount of work is proportional to the number of function evaluations, so that in general we have $W \sim f(n)/\delta t$, where n is the number of points and f gives the work done at each time step. Evidently $n \sim 1/h^2$; of course the exact constant of proportionality depends on the geometrical properties of the mesh, but for our purposes this is not essential. Ignoring the book-keeping the work is essentially one matrix vector multiplication for second order schemes and two for fourth order schemes. The matrices being sparse f is a linear function and the constant of proportionality can be taken as the number of non zero elements of one row of the matrix. With these conventions we get for the scheme (2.4) and (4.2)

$$\begin{aligned} W_2 &\sim \frac{5}{\delta t\, h_2^2} \\ W_4 &\sim \frac{14}{\delta t\, h_4^2} \end{aligned} \tag{4.4}$$

where the subscript 2 refers to second order scheme and 4 to fourth order scheme.

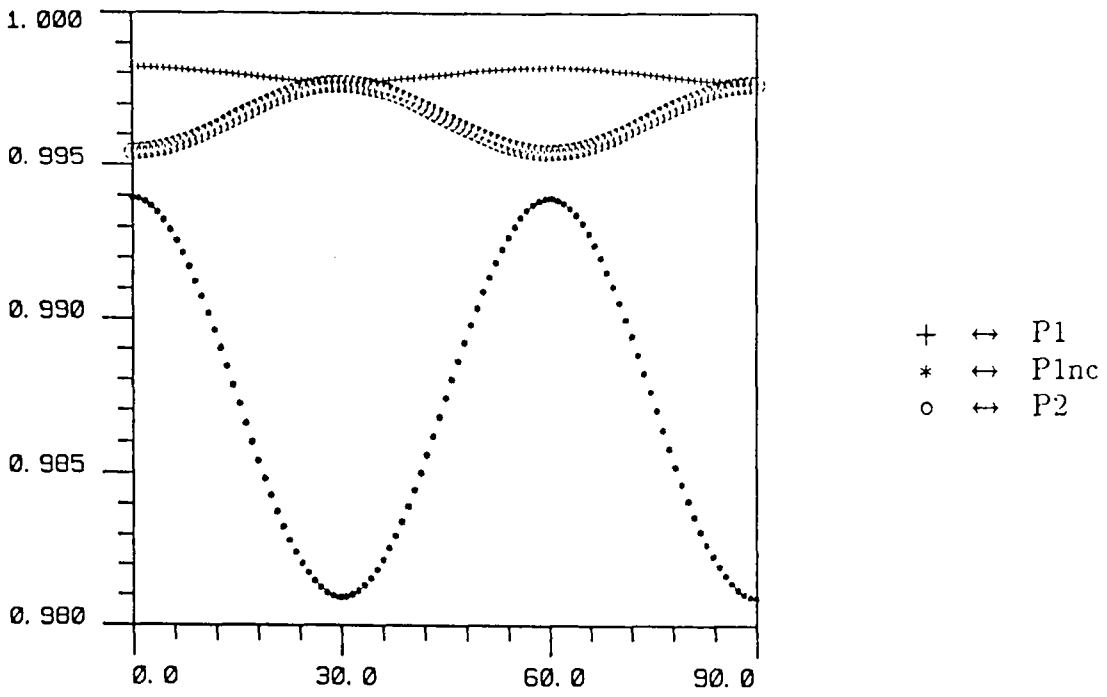


Figure 4.5: Comparison of different methods; $N = 5$ and $\alpha = 0.7$

Now using the best possible α for both methods we see that the work is about the same if

$$h_4 = \left(\frac{14}{5 \cdot 0.79\sqrt{2}}\right)^{1/3} h_2 \simeq 1.36 h_2$$

In figures 4.7, 4.8 and 4.9 we have plotted the phase speeds of these methods using the above scaling with $N = 3$, $N = 6$ and $N = 10$ respectively. The superiority of the fourth order scheme is quite clear and in addition we see that if a very accurate solution is required the advantage of the fourth order scheme is even greater.

Of course this approach does not take into account the storage requirements: sometimes it is more important to keep the size of the problem reasonable than to minimise the CPU time. However, recalling that compensation does not need any extra storage, the above crude analysis can be regarded as satisfactory.

Finally we remark that the compensation has no effect on the direction of the group speed, as can easily be verified.

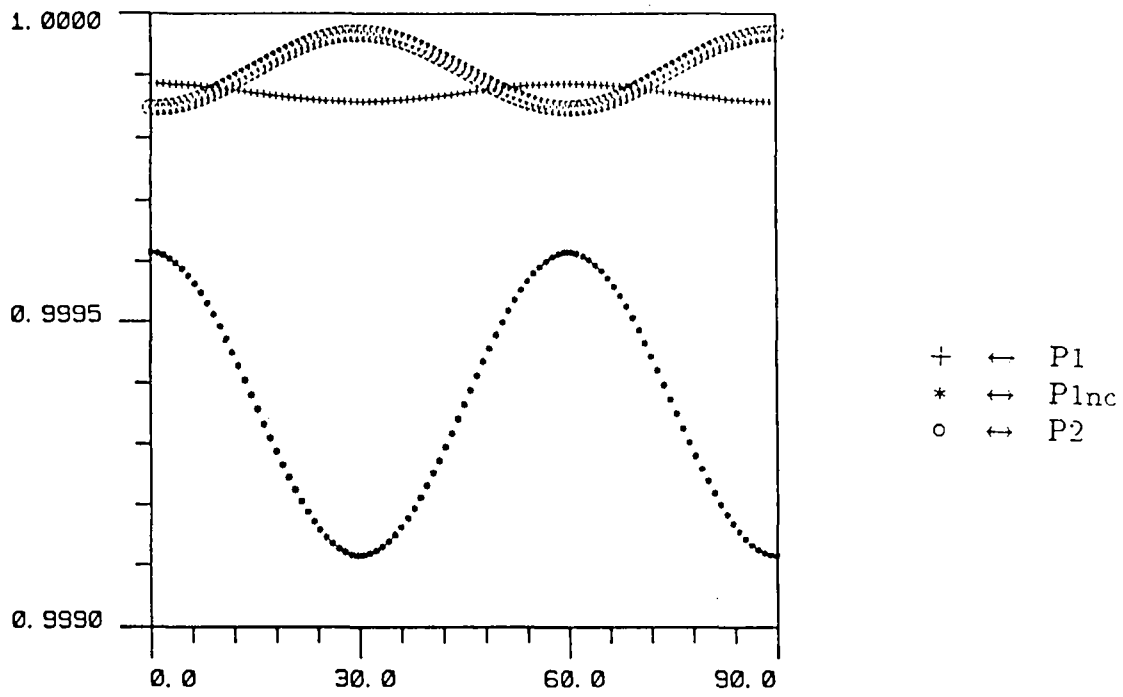


Figure 4.6: Comparison of different methods; $N = 10$ and $\alpha = 0.7$

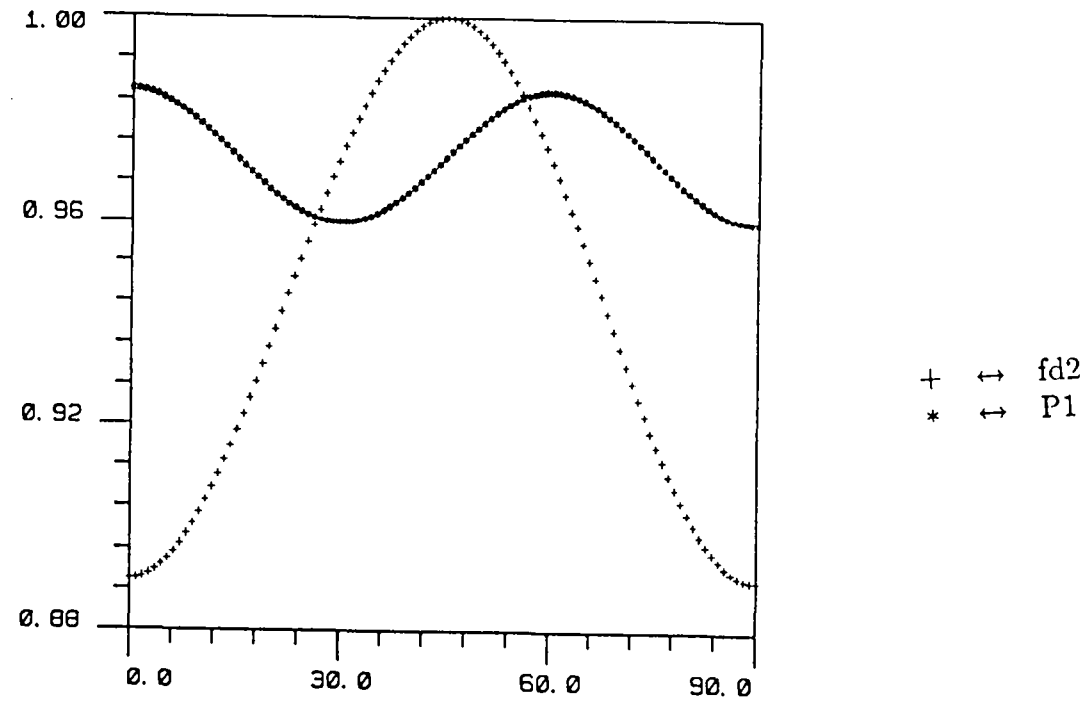


Figure 4.7: Comparison of second and fourth order methods with scaling; $N = 3$.

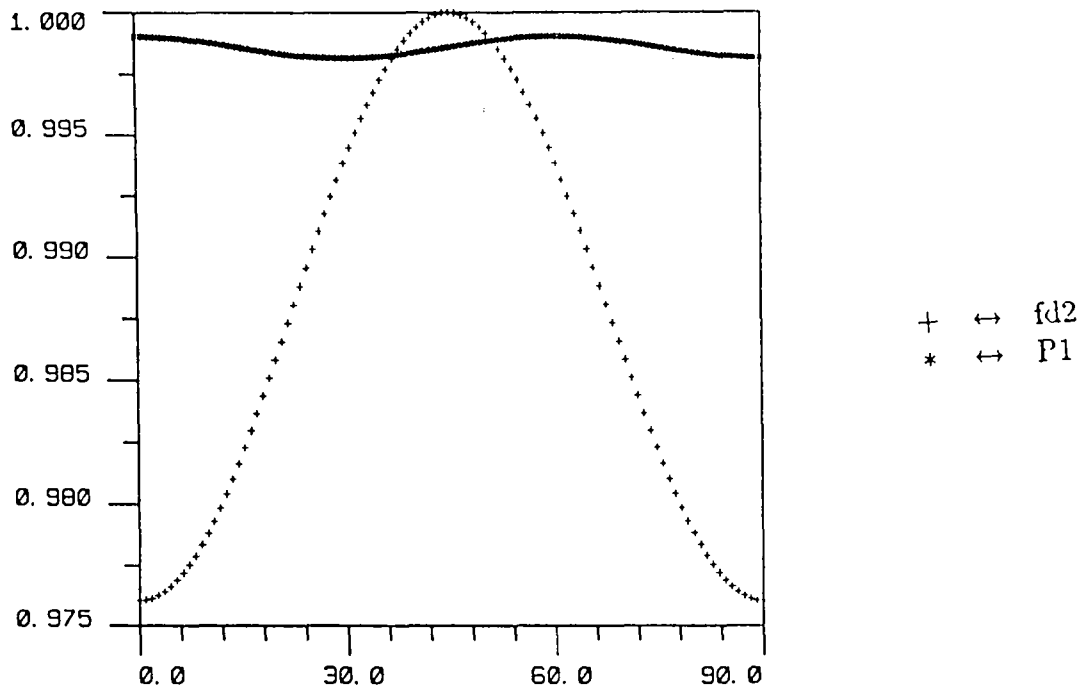


Figure 4.8: Comparison of second and fourth order methods with scaling; $N = 6$.

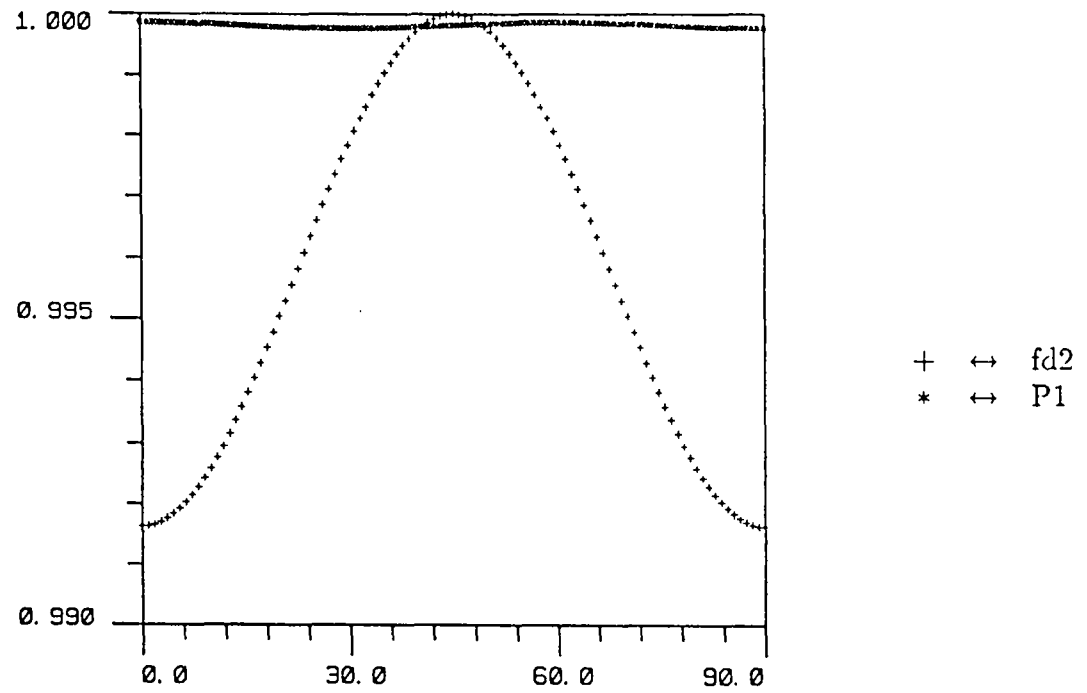


Figure 4.9: Comparison of second and fourth order methods with scaling; $N = 10$.

5 Three Dimensional Case

5.1 A Finite Difference Scheme

We have seen in the two dimensional case that it was essential for the compensation to work that the second order terms of the error are isotropic. The natural solution was to use equilateral triangles. In three dimensions there is no such easy solution (the space cannot be filled with regular tetrahedra). Using the finite differences one can however construct the following rather simple scheme to approximate the laplacian.

$$\begin{aligned} T_\Delta u \;=\; & \frac{1}{6}(24u(x,y,z) - 2u(x+h,y,z) - 2u(x-h,y,z) \\ & -2u(x,y+h,z) - 2u(x,y-h,z) - 2u(x,y,z+h) - 2u(x,y,z-h) \\ & -u(x+h,y+h,z) - u(x-h,y-h,z) - u(x-h,y+h,z) \\ & -u(x+h,y-h,z) - u(x+h,y,z+h) - u(x-h,y,z-h) \\ & -u(x-h,y,z+h) - u(x+h,y,z-h) - u(x,y+h,z+h) \\ & -u(x,y-h,z-h) - u(x,y-h,z+h) - u(x,y+h,z-h)) \end{aligned} \tag{5.1}$$

With the usual time discretisation we then have the method

$$\frac{u^{n+1} - 2u^n + u^{n-1}}{\delta t^2} + \frac{1}{h^2} T_\Delta u^n = 0$$

The corresponding dispersion relation is

$$\begin{aligned} & \frac{4}{\delta t^2}\sin^2(\omega\delta t/2) = \frac{2}{3h^2}(2\sin^2(a_1/2) + 2\sin^2(a_2/2) + 2\sin^2(a_3/2) + \\ & \sin^2((a_1+a_2)/2) + \sin^2((a_1-a_2)/2) + \sin^2((a_1+a_3)/2) + \\ & \sin^2((a_1-a_3)/2) + \sin^2((a_3+a_2)/2) + \sin^2((a_3-a_2)/2)) \end{aligned} \tag{5.2}$$

The Taylor's series gives us

$$\omega^2 = |k|^2(1 + (\delta t^2 - h^2)|k|^2/12 + O(h^4 + \delta t^4))$$

So the optimal time step would be $\delta t = h$. Let us denote by f_Δ the right hand side of (5.2) multiplied by h^2. Again with a little differential calculus one finds that the gradient of f_Δ vanishes when the following equations are satisfied

$$\begin{aligned} \sin a_1(1 + \cos a_2 + \cos a_3) &= 0 \\ \sin a_2(1 + \cos a_1 + \cos a_3) &= 0 \\ \sin a_3(1 + \cos a_2 + \cos a_1) &= 0 \end{aligned}$$

Then considering the different possibilities we see that

$$\max_{a \in \mathbb{R}^3} f_\Delta(a) = 16/3$$

Combining this with the stability condition

$$\alpha^2 f_\Delta \leq 4$$

gives us the limit

$$\alpha \leq \sqrt{3}/2 \simeq 0.87$$

As usual the optimal time step is unstable. We apply again the idea of compensation; the method and the corresponding dispersion relation then become

$$\frac{u^{n+1} - 2u^n + u^{n-1}}{\delta t^2} + \frac{1}{h^2}\left(T_\Delta u^n + \frac{h^2 - \delta t^2}{12h^2} T_\Delta^2 u^n\right) = 0$$

$$\frac{4}{\delta t^2}\sin^2(\omega \delta t/2) = \frac{1}{h^2}\left(f_\Delta + \frac{h^2 - \delta t^2}{12h^2} f_\Delta^2\right) \tag{5.3}$$

Then one obtains the stability condition (as before only the small values of α interest us so the coefficient before the f_Δ^2 term is always positive)

$$\alpha^2\left(f_\Delta + \frac{1 - \alpha^2}{12} f_\Delta^2\right) \leq 4 \tag{5.4}$$

Substituting the maximum value of f_Δ we have the inequality

$$16\alpha^4 - 52\alpha^2 + 27 \geq 0$$

Solving this we finally get

$$\alpha \leq \sqrt{\frac{13 - \sqrt{61}}{8}} \simeq 0.80 \tag{5.5}$$

Here also we notice that the above condition is not much stricter than without compensation.

Finally we note that instead of the operator T_Δ we could have considered a slightly more general operator which uses also the values $u(x \pm h, y \pm h, z \pm h)$. Then demanding that the second order error term be isotropic we get the following family of operators depending on one arbitrary parameter.

$$\begin{aligned}
\tilde{T}_\Delta u \;=\; & \frac{1}{6}(\tilde{d}u(x,y,z) - \tilde{a}(u(x+h,y,z) + u(x-h,y,z) + \\
& u(x,y+h,z) + u(x,y-h,z) + u(x,y,z+h) + u(x,y,z-h)) \\
& -\tilde{b}(u(x+h,y+h,z) + u(x-h,y-h,z) + u(x-h,y+h,z) + \\
& u(x+h,y-h,z) + u(x+h,y,z+h) + u(x-h,y,z-h) +
\end{aligned}$$

$$
\begin{aligned}
&u(x-h,y,z+h)+u(x+h,y,z-h)+u(x,y+h,z+h)+\\
&+u(x,y-h,z-h)+u(x,y-h,z+h)+u(x,y+h,z-h))\\
&-\tilde{c}(u(x+h,y+h,z+h)+u(x-h,y-h,z-h)+\\
&u(x-h,y+h,z+h)+u(x+h,y-h,z-h)+\\
&u(x+h,y-h,z+h)+u(x-h,y+h,z-h)+\\
&u(x-h,y-h,z+h)+u(x+h,y+h,z-h)))
\end{aligned}
$$

where the coefficients are given by

$$
\begin{aligned}
\tilde{d} &= 24+48\tilde{c}\\
\tilde{a} &= 2+24\tilde{c}\\
\tilde{b} &= 1-12\tilde{c}
\end{aligned}
$$

Taking $\tilde{c}=0$ gives T_Δ. Now one might think that by choosing a good value of $\tilde{c}$, one could improve accuracy and/or stability. However, this is not possible. First, calculating the dispersion relation

$$
\frac{4}{\delta t^2}\sin^2(\omega\delta t/2)=\tilde{f}_\Delta/h^2
$$

we see that $\tilde{f}_\Delta(\pi,\pi,0)=16/3$ for any $\tilde{c}$ so the stability condition cannot be improved. As regards the accuracy, we first calculate the fourth order error term, which gives

$$
\begin{aligned}
&(-4+5\alpha^2+\alpha^4+2(a_1^2+a_2^2)-14a_1^2a_2^2-2(a_1^4+a_2^4)\\
&+12(a_1^2a_2^4+a_1^4a_2^2)+90\tilde{c}a_1^2a_2^2(1-a_1^2-a_2^2))/360
\end{aligned}
$$

Here we have used $a_3^2=1-a_1^2-a_2^2$. Now it is straightforward to verify that when $\tilde{c}=0$ the maximum error is attained at $(a_1,a_2,a_3)=(0,0,1)$. Then referring to the above expression we see that the error in this direction does not depend on $\tilde{c}$. This shows that the accuracy cannot be improved. Consequently taking $\tilde{c}=0$ is the best choice because it reduces the amount of work without having any undesirable effects on stability or accuracy.

5.2 Tiling the Space and P1 Elements

We remarked above that the three dimensional space cannot be tiled with regular tetrahedra. However, in figure 5.1 there is an interesting tiling with 'almost regular' tetrahedra (for more details and other tilings by tetrahedra, see [GO] and [GS]). In particular all vertices are equivalent so there are no parasitic waves. Now there is an infinite family of different tilings, because the ratio a/e (in the notation of figure 5.1) can be chosen arbitrarily. The most natural choice seems to be choosing the

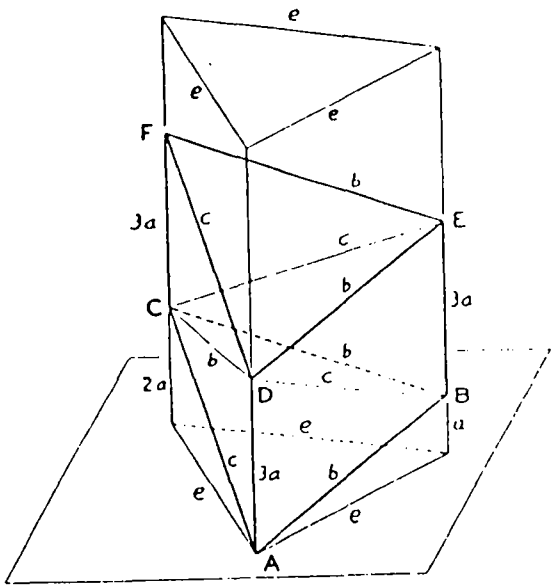


Figure 5.1: Tiling of the space by congruent tetrahedra

ratio such that $3a = b$, which implies that all the four faces are congruent isosceles triangles. Let us denote the shorter edge by h (the longer edge is then $2h/\sqrt{3}$). Then define the following vectors.

$$\begin{aligned} h_1 &= h(\sqrt{8}/3 \quad 0 \quad 1/3) \\ h_2 &= h(\sqrt{2}/3 \quad \sqrt{6}/3 \quad -1/3) \\ h_3 &= h(-\sqrt{2}/3 \quad \sqrt{6}/3 \quad 1/3) \\ h_4 &= h(0 \quad 0 \quad 1) \end{aligned}$$

After the usual finite element calculations using P1 elements we find that the method can be interpreted as the following difference scheme

$$\frac{u^{n+1} - 2u^n + u^{n-1}}{\delta t^2} + \frac{1}{h^2}\check{T}_\Delta u^n = 0 \tag{5.6}$$

where the operator $\check{T}_\Delta$ is defined by

$$\check{T}_\Delta u = \frac{3}{4}\left(8u(r) - \sum_{i=1}^{4} u(r \pm h_i)\right)$$

Note in particular that all the four directions h_i are equivalent, which would not have been the case with other choices of the parameter a/e. This leads to the following dispersion relation.

$$\frac{4}{\delta t^2}\sin^2(\omega\delta t/2) = \frac{3}{h^2}\sum_{i=1}^{4}\sin^2(k\cdot\bar{h}_i/2) \tag{5.7}$$

Expanding the above equation gives

$$\omega^2 = |k|^2(1+(9\delta t^2 - g(k)h^2)|k|^2/108 + O(h^4+\delta t^4))$$

where g is given by

$$g(k) = (6k_1^4 + 6k_2^4 + 7k_3^4 + 12k_1^2k_2^2 + 6k_1^2k_3^2 + 6k_3^2k_2^2 + 4\sqrt{2}k_1^3k_2 + 12\sqrt{2}k_1k_2^2k_3)/|k|^4$$

Now it is straight forward to calculate that $3 \le g \le 7$, so part of the error can be compensated by the following scheme.

$$\frac{u^{n+1}-2u^n+u^{n-1}}{\delta t^2} + \frac{1}{h^2}\left(\check{T}_\Delta u^n + \frac{5h^2-9\delta t^2}{108h^2}\check{T}_\Delta^2 u^n\right) = 0 \tag{5.8}$$

The final dispersion relation is then of the form

$$\omega^2 = |k|^2(1+\check{g}(k)h^2|k|^2/108 + O(h^4+\delta t^4))$$

where $|\check{g}| \le 2$, so that the maximum error in the space discretisation is reduced by a factor of $7/2$. Let us remark that it is easy to see that we cannot compensate the error totally with any value of a/e. In particular if we denote the right hand side of (5.7) by f_{P1}, then $\partial^4 f_{P1}/\partial^3 k_1 \partial k_3 \neq 0$ (at the origin) for all $a \neq 0$, which clearly implies that compensation is impossible (see proposition 1 and conditions (C1)-(C5)).

As regards the stability condition, it is particularly easy to calculate: it is sufficient to note that when $k = (\,0 \quad 0 \quad 3\pi/h\,)$, the right hand side of (5.7) is $12/h^2$ which evidently is the global maximum. So the stability condition for (5.6) is

$$\alpha \le 1/\sqrt{3} \simeq 0.58$$

The stability inequality for the scheme (5.8) is then

$$\alpha^2\left(12 + \frac{5-9\alpha^2}{108}144\right) \le 4$$

Solving this finally gives

$$\alpha \le \frac{\sqrt{7-\sqrt{22}}}{3} \simeq 0.51$$

5.3 Comparing the Different Schemes

Next we will compare the above methods like in two dimensional case. We represent the wave vector in spherical coordinates in the following way (scaling $|k|$ to be one as usual)

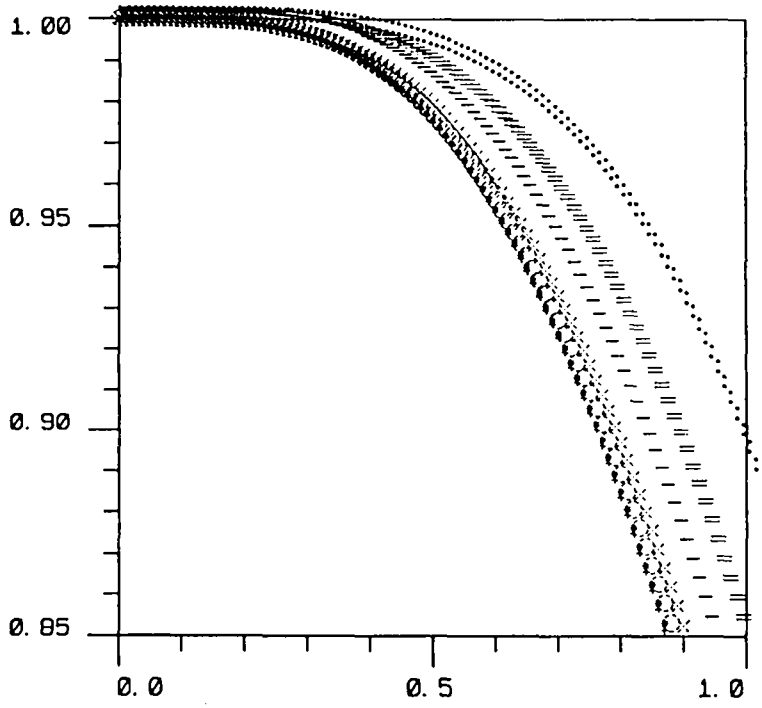
$$\begin{aligned} k_1 &= \cos\varphi\cos\theta \\ k_2 &= \sin\varphi\cos\theta \\ k_3 &= \sin\theta \end{aligned}$$

In figure 5.2 we see the phase speed as a function of $2/N$ for the method (5.3) with different values of α. As usual, the accuracy improves when α grows. In figure 5.3 there is a corresponding picture for the scheme (5.8). The directions chosen are typical: other values of φ and θ give similar results. Finally in figure 5.4 we have plotted the two methods with their best value of α (that is, at the stability limit).

Next, doing a similar crude operation count as in (4.4) and then using the corresponding scaling, there is in figure 5.5 the phase speed as a function of θ for the two methods for different values of parameters. Similarly, in figure 5.6 we have compared the ordinary second order scheme ((2.4) in three dimensions) to the fourth order scheme, scaling the work to be the same. The superiority of the scheme (5.3) is quite obvious in both cases.

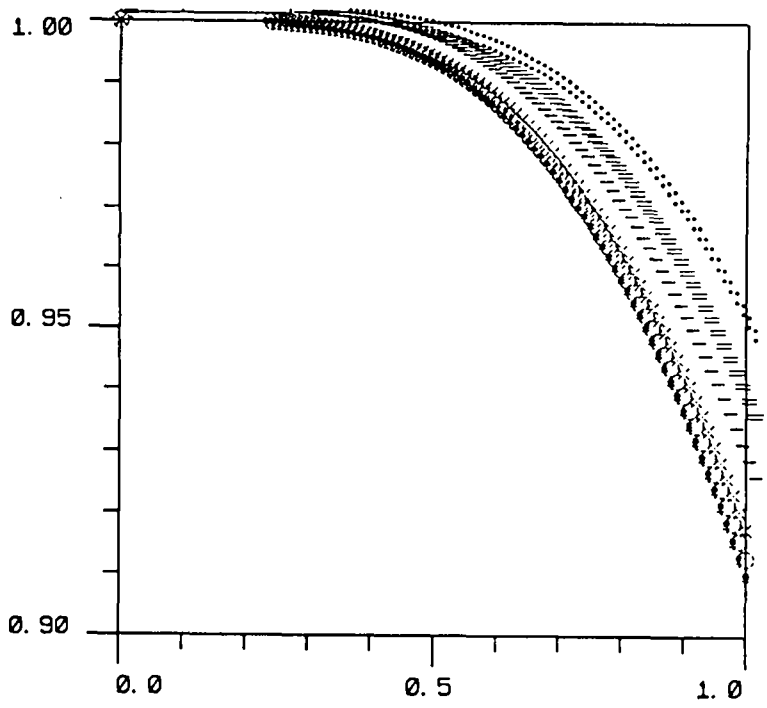
5.4 Direction of Propagation

Looking at the error in the direction of the group speed we see that the situation is exactly like in the two dimensional case: proposition 1 tells us that with or without compensation we have $\beta = O(h^4)$ for the method (5.3). In figure 5.7 and 5.8 we show β with different values of the parameters. It is seen that β is very small already for quite small values of N.



$+ \leftrightarrow \alpha = 0$
$* \leftrightarrow \alpha = 0.80/6$
$o \leftrightarrow \alpha = 2 \cdot 0.80/6$
$\times \leftrightarrow \alpha = 3 \cdot 0.80/6$
$- \leftrightarrow \alpha = 4 \cdot 0.80/6$
$= \leftrightarrow \alpha = 5 \cdot 0.80/6$
$: \leftrightarrow \alpha = 0.80$

Figure 5.2: Phase speed; $\varphi = \theta = 45$



$+ \leftrightarrow \alpha = 0$
$* \leftrightarrow \alpha = 0.51/6$
$o \leftrightarrow \alpha = 2 \cdot 0.51/6$
$\times \leftrightarrow \alpha = 3 \cdot 0.51/6$
$- \leftrightarrow \alpha = 4 \cdot 0.51/6$
$= \leftrightarrow \alpha = 5 \cdot 0.51/6$
$: \leftrightarrow \alpha = 0.51$

Figure 5.3: Phase speed; $\varphi = \theta = 45$

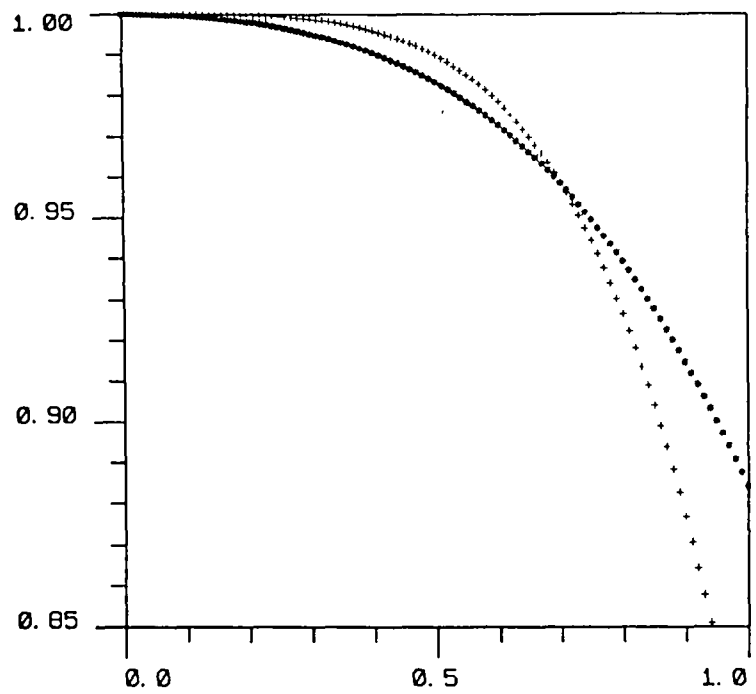


Figure 5.4: Phase speed; $\varphi = \theta = 0$

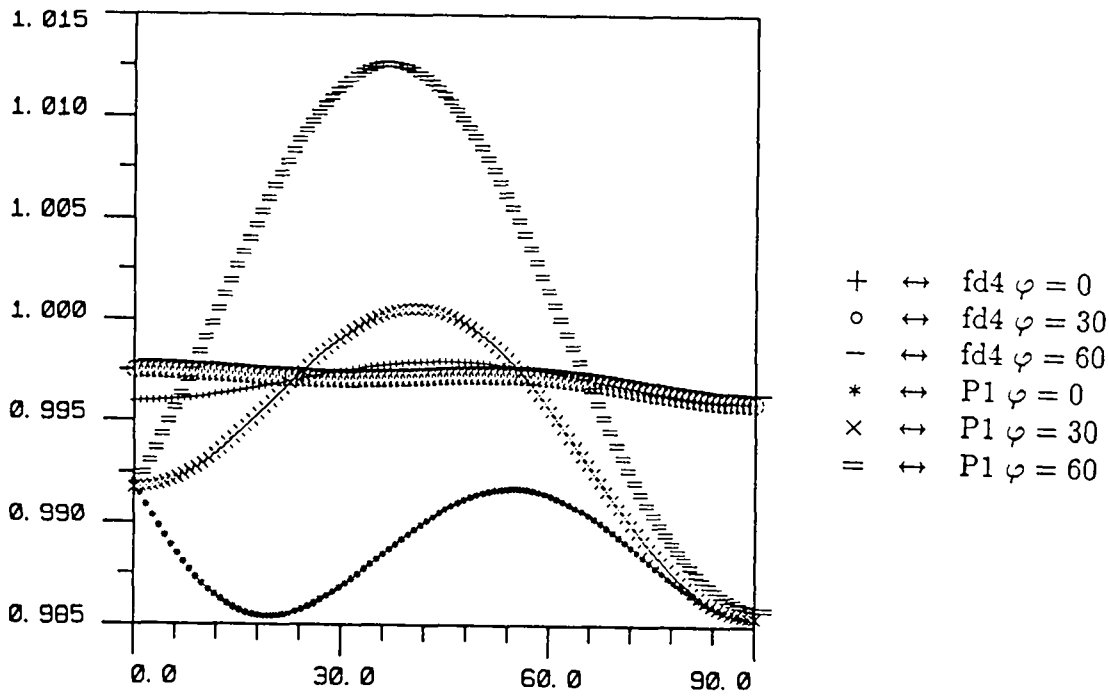


Figure 5.5: Phase speed, $N = 5$

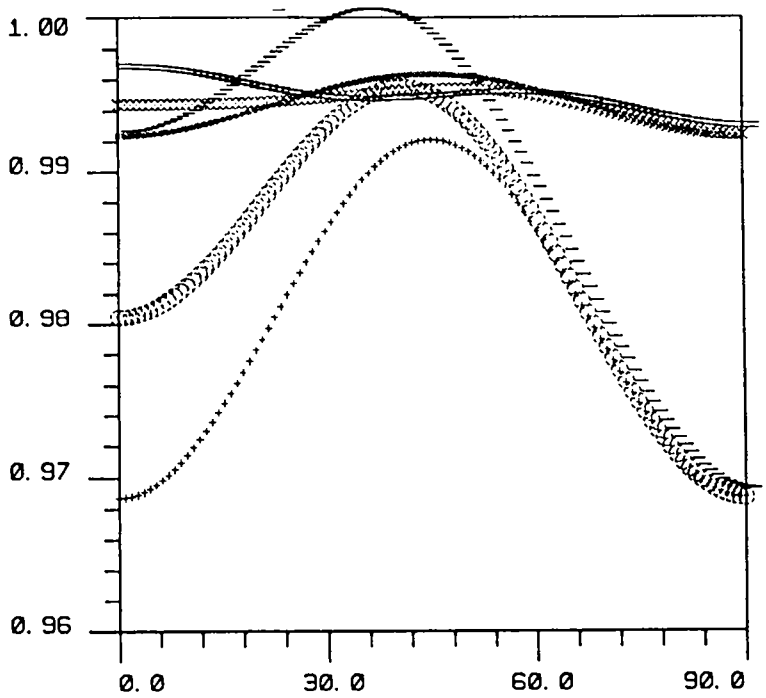


Figure 5.6: Phase speed, $N = 6$

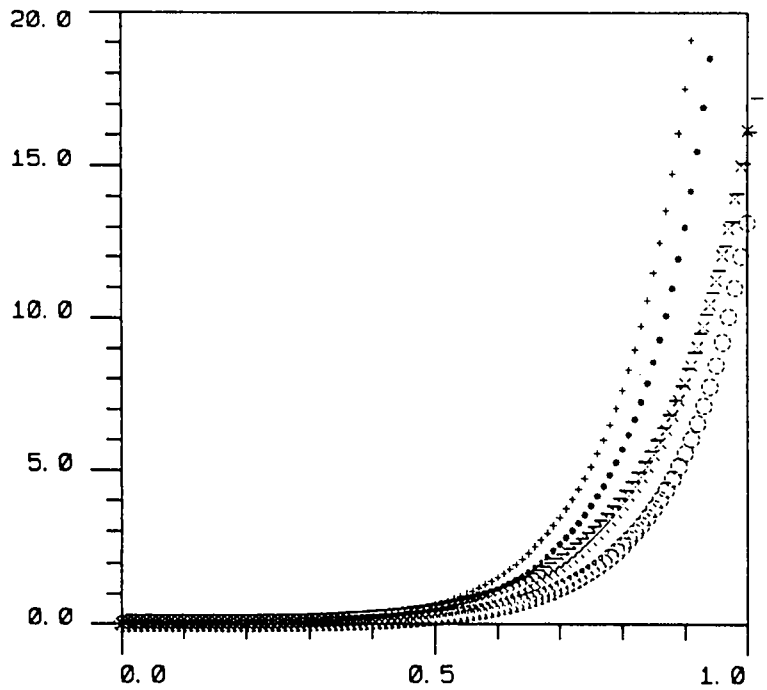


Figure 5.7: β as a function of $2/N$, $\varphi = 20$

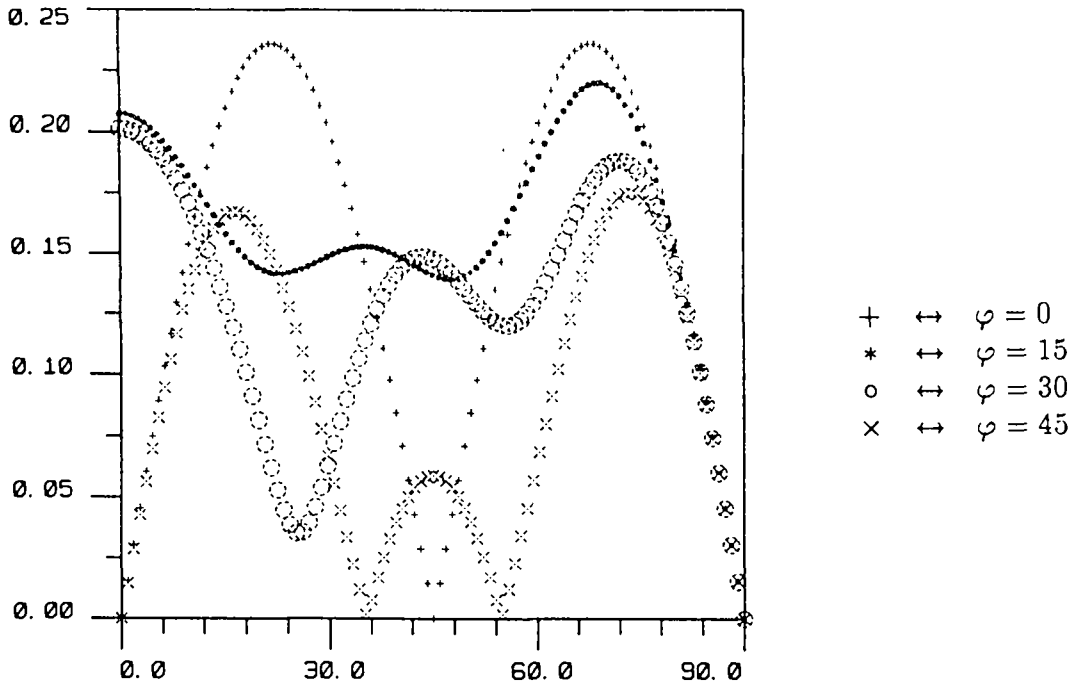


Figure 5.8: β as a function of θ, $N = 5$

6 Maxwell's Equations

6.1 Background

Next we generalise the above ideas to the Maxwell's equations. Considering a domain in space free of sources we can write the equations in the following form.

$$\begin{aligned} E_t - \nabla \times B &= 0 \\ H_t + \nabla \times D &= 0 \\ \nabla \cdot D &= 0 \\ \nabla \cdot B &= 0 \end{aligned} \tag{6.1}$$

For the physical interpretation of different unknowns we refer to [LL] and [KO]. We just remark that each of the above functions are vector valued and that last two equations are really consequences of the first two. This leaves us with six independent scalar equations with twelve unknowns so we need the so called constitutive relations which determine the dependence of E and H of D and B. We will consider only isotropic media and so the constitutive equations are simply

$$\begin{aligned} D &= \epsilon E \\ B &= \mu H \end{aligned} \tag{6.2}$$

The coefficients ϵ and μ are called permittivity and permeability. For simplicity of notation we take these constants to be one so our fundamental equations become

$$\begin{aligned} D_t - \nabla \times B &= 0 \\ B_t + \nabla \times D &= 0 \end{aligned} \tag{6.3}$$

With this choice the speed of light is one; with physical parameters it would be $c = 1/\sqrt{\epsilon\mu}$. These equations admit the plane wave solutions of the form

$$\begin{aligned} B &= B_0 e^{i(k\cdot r - \omega t)} \\ D &= D_0 e^{i(k\cdot r - \omega t)} \end{aligned} \tag{6.4}$$

where B_0 and D_0 are constant vectors. Substituting the above plane waves into (6.3) we find

$$\begin{aligned} \omega D_0 &= -k \times B_0 \\ \omega B_0 &= k \times D_0 \end{aligned} \tag{6.5}$$

This is really an eigenvalue problem which is seen more clearly when one associates to the operator $k\times$ a 3×3 matrix as follows.

$$K = k\times = \begin{pmatrix} 0 & -k_3 & k_2 \\ k_3 & 0 & -k_1 \\ -k_2 & k_1 & 0 \end{pmatrix} \tag{6.6}$$

Note that K is antisymmetric and so its eigenvalues are purely imaginary and by a simple calculation they are seen to be 0 and $\pm i\,|k|$. Ignoring the eigenvalue zero we then get the following dispersion relation

$$\omega^2 = |k|^2$$

As regards the eigenvectors we obtain immediately that

$$\begin{aligned} k \cdot B_0 &= 0 \\ k \cdot D_0 &= 0 \end{aligned} \tag{6.7}$$

So when B and D are of the form (6.4) they are orthogonal to the direction of propagation (and to each other). This is also true in more general situations where k might not be orthogonal to E or H, see [KO]. The above conditions also clearly imply that B and D as given in (6.4) satisfy the last two equations of (6.1).

6.2 Basic Second Order Scheme

We start with some notations and define the following difference operators. Here ψ is a function from $\mathbb{R}^1$ to $\mathbb{R}^3$ and ψ_i refers obviously to its i'th component.

$$V_x\psi_i = \frac{\psi_i(x+h,y,z,t) - \psi_i(x-h,y,z,t)}{2}$$

Exactly in the same way we define the difference operators in y and z directions. Then to approximate $\nabla\times$ we use

$$\bar{T}_\times\psi = \begin{pmatrix} 0 & -V_z & V_y \\ V_z & 0 & -V_x \\ -V_y & V_x & 0 \end{pmatrix}\begin{pmatrix}\psi_1\\ \psi_2\\ \psi_3\end{pmatrix} = \begin{pmatrix} V_y\psi_3 - V_z\psi_2 \\ V_z\psi_1 - V_x\psi_3 \\ V_x\psi_2 - V_y\psi_1\end{pmatrix}$$

We thus have the following numerical method for solving (6.3).

$$\begin{aligned} \frac{D^{n+1} - D^{n-1}}{2\delta t} - \frac{1}{h}\bar{T}_\times B^n &= 0 \\ \frac{B^{n+1} - B^{n-1}}{2\delta t} + \frac{1}{h}\bar{T}_\times D^n &= 0 \end{aligned}$$

Now to analyze the scheme we use the plane waves as before. First note that

$$V_x e^{i(k\cdot r - \omega t)} = i\sin(k_1 h)e^{i(k\cdot r-\omega t)}$$

So instead of the operator K in (6.6) we have

$$\bar{\mathcal{K}} = \begin{pmatrix} 0 & -\sin a_3 & \sin a_2 \\ \sin a_3 & 0 & -\sin a_1 \\ -\sin a_2 & \sin a_1 & 0 \end{pmatrix}$$

where as before we have denoted $k_i h$ by a_i . Our eigenvalue problem then becomes

$$\begin{aligned} \frac{\sin(\omega\delta t)}{\delta t} D_0 &= -\frac{1}{h}\bar{\mathcal{K}}B_0 \\ \frac{\sin(\omega\delta t)}{\delta t} B_0 &= \frac{1}{h}\bar{\mathcal{K}}D_0 \end{aligned}$$

Solving this is as easy as solving (6.5), so using the notation $\sin a = (\sin a_1 \quad \sin a_2 \quad \sin a_3)$ we get the dispersion relation

$$\frac{1}{\delta t^2}\sin^2(\omega\delta t) = \frac{1}{h^2}|\sin a|^2 \tag{6.8}$$

Denoting the (non zero) eigenvalues of $\bar{K}$ by $\pm i\bar{\nu}$, we see that the right hand side of (6.8) is just $\bar{\nu}^2/h^2$. In addition, instead of the original orthogonality relations (6.7) we have

$$\begin{aligned} \sin a \cdot B_0 &= 0 \\ \sin a \cdot D_0 &= 0 \end{aligned}$$

From (6.8) we get the stability condition

$$\alpha^2 |\sin a|^2 \leq 1$$

The components of k being independent this gives

$$\alpha \leq 1/\sqrt{3} \simeq 0.58$$

We can improve the stability using the following idea. To 'simulate' a small timestep and keep the discretisation centered we could calculate D (resp. B) at instants $n\,\delta t$ (resp. $(n+1/2)\delta t$). This approach gives the method

$$\begin{aligned} \frac{D^{n+1} - D^n}{\delta t} - \frac{1}{h}\bar{T}_\times B^{n+1/2} &= 0 \\ \frac{B^{n+1/2} - B^{n-1/2}}{\delta t} + \frac{1}{h}\bar{T}_\times D^n &= 0 \end{aligned}$$

Making the same plane wave analysis as above we notice that everything remains the same, except that one replaces δt by $\delta t/2$ everywhere. So the stability condition becomes

$$\alpha \leq 2/\sqrt{3} \simeq 1.15$$

This seems to be a good argument for preferring this kind of time discretisation, so to avoid unnecessary repetition we consider from now on only the case where the fields B and D are given at different time levels. Finally, using the Taylor's expansion to calculate the asymptotic behavior of the new dispersion relation we obtain

$$\omega^2 = |k|^2(1 + (|k|^2\delta t^2 - 4h^2\frac{k_1^4 + k_2^4 + k_3^4}{|k|^2})/12 + O(h^4 + \delta t^4))$$

We conclude that the scheme is of the second order and that it is clearly anisotropic.

6.3 Fourth Order Scheme

To get a fourth order scheme in time we have exactly the same difficulty as with the wave equation: to avoid dissipation the time discretisation has to be centered, and to have a stable scheme it has to be of second order. To circumvent this problem we

use the same modified equation approach as before. To this end, let us write down the truncation error in time discretisation.

$$\begin{aligned}
\frac{D^{n+1}-D^n}{\delta t} &= \partial_t D^{n+1/2} + \frac{\delta t^2}{24}\partial_t^3 D^{n+1/2} + O(\delta t^4) \\
&= \partial_t D^{n+1/2} - \frac{\delta t^2}{24}(\nabla\times)^3 B^{n+1/2} + O(\delta t^4) \\
\frac{B^{n+1/2}-B^{n-1/2}}{\delta t} &= \partial_t B^n + \frac{\delta t^2}{24}\partial_t^3 B^n + O(\delta t^4) \\
&= \partial_t B^n + \frac{\delta t^2}{24}(\nabla\times)^3 D^n + O(\delta t^4)
\end{aligned}$$

Now if we add the second order error term to the equations (6.3), discretise it suitably and use a fourth order scheme for $\nabla\times$ then we get a fourth order scheme. More generally if we use a second order discretisation for $\nabla\times$ and the error is of the same form as above, then it can also be compensated exactly in the same way. To analyse this possibility we introduce the following rather complicated difference operator (we omit the time argument of ψ).

$$\begin{aligned}
S_x^c\psi_i &= \frac{1}{12}(-\psi_i(x+2h,y,z) + 8\psi_i(x+h,y,z) - 8\psi_i(x-h,y,z) \\
&\quad +\psi_i(x-2h,y,z)) + \frac{c}{2h}(\psi_i(x+2h,y,z) - 6\psi_i(x+h,y,z) \\
&\quad +6\psi_i(x-h,y,z) - \psi_i(x-2h,y,z)) + \frac{c}{2h}(\psi_i(x+h,y+h,z) \\
&\quad -\psi_i(x-h,y-h,z) + \psi_i(x+h,y-h,z) - \psi_i(x-h,y+h,z) \\
&\quad +\psi_i(x+h,y,z+h) - \psi_i(x-h,y,z-h) \\
&\quad +\psi_i(x+h,y,z-h) - \psi_i(x-h,y,z+h))
\end{aligned} \tag{6.9}$$

In particular, choosing $c = 0$ gives the usual fourth order difference operator. Of course the y and z directions are handled similarly. Then to approximate $\nabla\times$ we naturally define

$$T_c\psi = \begin{pmatrix} 0 & -S_z^c & S_y^c \\ S_z^c & 0 & -S_x^c \\ -S_y^c & S_x^c & 0 \end{pmatrix}\begin{pmatrix}\psi_1\\ \psi_2\\ \psi_3\end{pmatrix} = \begin{pmatrix} S_y^c\psi_3 - S_z^c\psi_2 \\ S_z^c\psi_1 - S_x^c\psi_3 \\ S_x^c\psi_2 - S_y^c\psi_1 \end{pmatrix}$$

We have then the following result.

Proposition 1

$$\frac{1}{h}T_c\psi = \nabla\times\psi - ch^2(\nabla\times)^3\psi + O(h^4) \tag{6.10}$$

This is proved by first using Taylor's expansion in (6.9), which gives

$$\frac{1}{h}S_x^c\psi_i = \partial_x\psi_i + ch^2\partial_x\Delta\psi_i + O(h^4)$$

This implies that

$$\frac{1}{h}T_c\psi = \nabla\times\psi + ch^2\Delta\nabla\times\psi + O(h^4)$$

Now using the identity $\nabla\times\nabla\times\nabla\times\psi = -\Delta\nabla\times\psi$ we get (6.10).
Finally we have to discretise $\nabla\times\nabla\times\nabla\times$ in some way; however this is rather easy: we simply iterate the operator T_c to obtain

$$\frac{1}{h^3}T_c^3\psi = (\nabla\times)^3\psi + O(h^2)$$

This is only of the second order (if $c \neq 0$), but this does no harm because second order accuracy is sufficient, exactly for the same reasons as in the case of wave equation.
Before considering the full fourth order scheme, we first establish the stability properties of the following second order scheme.

$$\begin{aligned}\frac{D^{n+1}-D^n}{\delta t} - \frac{1}{h}T_cB^{n+1/2} &= 0\\ \frac{B^{n+1/2}-B^{n-1/2}}{\delta t} + \frac{1}{h}T_cD^n &= 0\end{aligned} \tag{6.11}$$

Using the plane wave analysis as before we first note that

$$S_x^c e^{i(k\cdot r-\omega t)} = \frac{i\sin a_1}{3}(4-\cos a_1 + 6c(\cos a_1+\cos a_2+\cos a_3-3))e^{i(k\cdot r-\omega t)}$$

Let us denote the expression multiplying the exponential in the right hand side of the above equation by $i\xi_{xc}$ (and similarly for $i\xi_{yc}$ and $i\xi_{zc}$). This gives us the operator

$$\mathcal{K}_c = \begin{pmatrix} 0 & -\xi_{zc} & \xi_{yc}\\ \xi_{zc} & 0 & -\xi_{xc}\\ -\xi_{yc} & \xi_{xc} & 0\end{pmatrix}$$

and the corresponding eigenvalue problem

$$\begin{aligned}\frac{2\sin(\omega\delta t/2)}{\delta t}D_0 &= -\frac{1}{h}\mathcal{K}_cB_0\\ \frac{2\sin(\omega\delta t/2)}{\delta t}B_0 &= \frac{1}{h}\mathcal{K}_cD_0\end{aligned}$$

Then proceeding as when deriving (6.8) we get

$$\frac{4}{\delta t^2}\sin^2(\omega\delta t/2) = \frac{1}{h^2}(\xi_{xc}^2+\xi_{yc}^2+\xi_{zc}^2)$$

We then define f_c to be the right hand side of the above equation multiplied by h^2, which gives

$$f_c(a) = \frac{1}{9}\sum_{i=1}^{3} \sin^2 a_i(4 - \cos a_i + 6c(-3 + \sum_{l=1}^{3} \cos a_l))^2$$

We remark that using the notation $\xi_c = (\,\xi_{xc} \quad \xi_{yc} \quad \xi_{zc}\,)$ we have $f_c = |\xi_c|^2$ Then choosing $c = 0$ gives the following stability condition.

$$\alpha^2 f_0 = \frac{\alpha^2}{9}\sum_{i=1}^{3} \sin^2 a_i(4 - \cos a_i)^2 \leq 4$$

A little differential calculus then shows that the maximum value of the left hand side is attained when $\cos a_i = 1 - \sqrt{6}/2$, so that the maximum of the f_0 is $M_0 = (8 + \sqrt{6})/4 \simeq 5.65$ (in what follows we shall denote the maximum value of f_c by M_c). The stability condition then becomes

$$\alpha \leq \frac{4}{\sqrt{3 + 8\sqrt{6}}} \simeq 0.84$$

The stability condition is rather reasonable, but maybe it could be improved by choosing another value of c. A little thought suggests that $c = 2/9$ might be a good candidate. The stability condition then becomes

$$\alpha^2 f_{2/9} = \frac{\alpha^2}{81}\sum_{i=1}^{3} \sin^2 a_i(\cos a_i + 4\sum_{l\neq i} \cos a_l)^2 \leq 4$$

Searching for the maximum of f_c is rather complicated, but numerically it can easily be calculated. In fact in figure 6.1 we have plotted M_c as a function of c and it clearly indicates that the minimum is somewhere near the value $c = 2/9$.

The numerical values of the point where $M_{2/9}$ is attained suggest that at the maximum $a_2 = 0$ and $a_3 = 0$. After this we easily calculate that the gradient vanishes at $\cos a_1 = 2 - 3/\sqrt{2}$, $a_2 = 0$ and $a_3 = 0$ (of course any other permutation of variables would do as well, and one could take π instead of zero). Consequently the maximum of the $f_{2/9}$ is $M_{2/9} = (16\sqrt{2} - 13)/12 \simeq 0.80$. This leads to the stability condition

$$\alpha \leq \frac{4\sqrt{3}}{\sqrt{16\sqrt{2} - 13}} \simeq 2.23$$

We then consider the fourth order scheme which can be written as follows.

$$\begin{aligned} &\frac{D^{n+1} - D^n}{\delta t} - \frac{1}{h}(T_c B^{n+1/2} - \beta_c T_c^3 B^{n+1/2}) = 0 \\ &\frac{B^{n+1/2} - B^{n-1/2}}{\delta t} + \frac{1}{h}(T_c D^n - \beta_c T_c^3 D^n) = 0 \end{aligned} \tag{6.12}$$

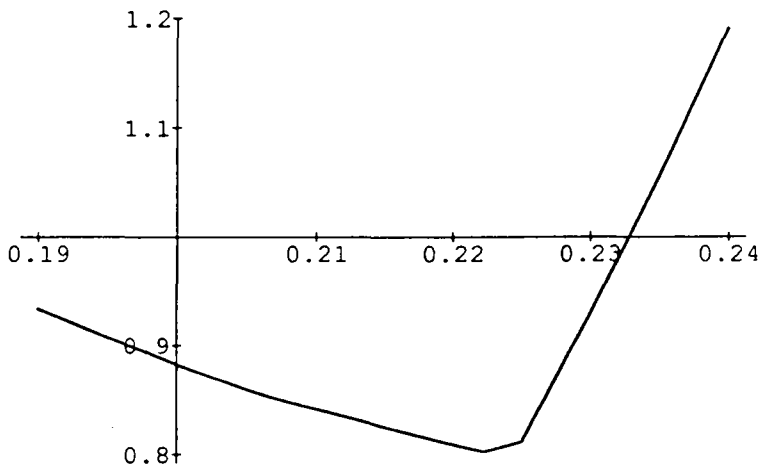


Figure 6.1: M_c as a function of c

where $\beta_c = (\delta t^2 - 24ch^2)/24h^2 = (\alpha^2 - 24c)/24$. Because of the special form of the correction terms (one just iterates T_c), the stability analysis is rather straightforward and proceeding as before we find the following eigenvalue problem.

$$\begin{aligned} \frac{2\sin(\omega\delta t/2)}{\delta t} D_0 &= -\frac{1}{h}(\mathcal{K}_c + \beta_c \mathcal{K}_c^3) B_0 \\ \frac{2\sin(\omega\delta t/2)}{\delta t} B_0 &= \frac{1}{h}(\mathcal{K}_c + \beta_c \mathcal{K}_c^3) D_0 \end{aligned}$$

Remembering that the (non zero) eigenvalues of $\mathcal{K}_c$ are $\pm i\,|\xi_c|$, then evidently the eigenvalues of $\mathcal{K}_c + \beta_c \mathcal{K}_c^3$ are $\pm i\,|\xi_c|\,(1 - \beta_c |\xi_c|^2)$. Then using f_c gives the dispersion relation

$$\frac{4}{\delta t^2} \sin^2(\omega\delta t/2) = \frac{f_c}{h^2}(1 - \beta_c f_c)^2 \tag{6.13}$$

Now maximising the right hand side is not so straightforward as before because β_c is positive for reasonable values of parameters. However, numerically we see that it is sufficient to consider the maximum value of f_c. This means that for each c the stability condition is given by

$$\alpha^2 M_c (1 + cM_c - \alpha^2 M_c/24)^2 - 4 \leq 0 \tag{6.14}$$

This is a third degree polynomial with respect to α^2 and so the above inequality can be solved exactly. However the expression is so complicated that we do not attempt to write it down. Instead we have plotted in figure 6.2 the left hand side of (6.14) for some values of c and curiously it happens that the stability limit does not change

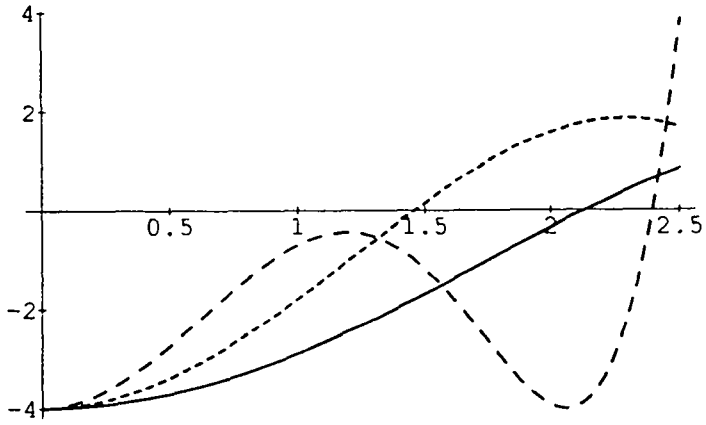


Figure 6.2: Right hand side of the stability inequality as a function of α with $c = 0$ (broken line with large dashes), $c = 0.1$ (broken line with small dashes) and $c = 0.207$ (unbroken line).

monotonically with respect to c. Also for the small values of c (including zero) the stability improves considerably compared to the corresponding second order scheme. Of course we would like to choose such a value of c that minimises the error. To this end we calculate the Taylor's expansion of (6.13) which gives

$$\omega = |k|(1 + g(\alpha, c, a_1, a_2, a_3)h^4|k|^4 + O(h^6 + \delta t^6)) \tag{6.15}$$

Now g is homogeneous of degree zero with respect to $a = (\, a_1 \quad a_2 \quad a_3\,)$ so we can choose $a_3 = \sqrt{1 - a_1^2 - a_2^2}$ and then plot g for some values of α and c. For small values of c the result looks like in figure 6.3 where we have taken $c = 0$ and $\alpha = 1$. Evidently the maximum error (the minimum value of g) occurs at $a_1 = a_2 = 0$ and then we have

$$g(\alpha, 0, 0, 0, 1) = -\frac{\alpha^4 + 64}{1920} \tag{6.16}$$

On the other hand when $c \simeq 0.2$ the global minimum of g is at $a_1 = a_2 = a_3 = 1/\sqrt{3}$ and the situation is like in the figure 6.4 (with $\alpha = 2.13$ and $c = 0.207$). The minimum value is then

$$g(\alpha, c, 1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}) = -\frac{\alpha^4 + 64/9 + 5760c^2 - 240\alpha^2 c - 160c}{1920}$$

Now it turns out that if one fixes c and then tries to minimise the absolute value of

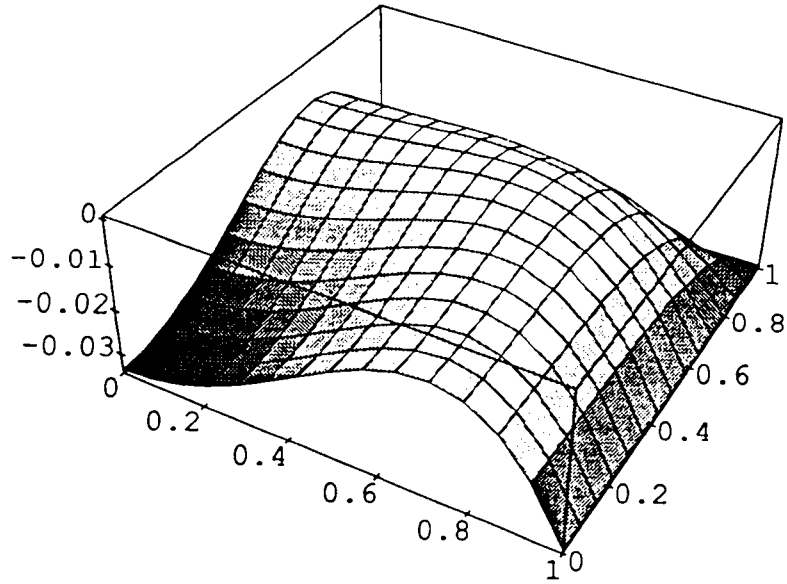


Figure 6.3: g as a function of a_1 and a_2; $c = 0$ and $\alpha = 1$.

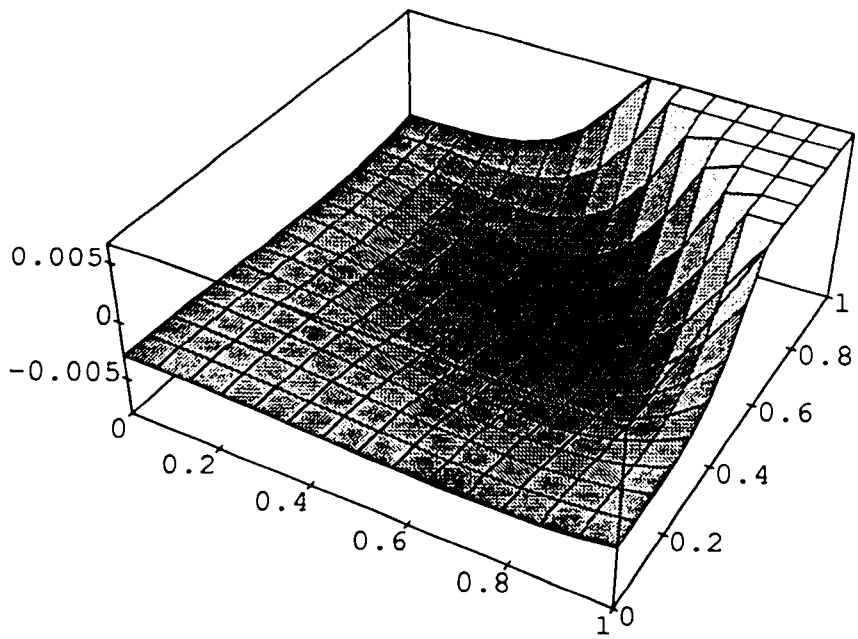


Figure 6.4: g as a function of a_1 and a_2; $c = 0.207$ and $\alpha = 2.13$.

the above expression with respect to α, then α at the minimum does not satisfy the stability condition. So the values used in figure 6.4 for α and c are merely numerically found values which should not be too far from the optimum. With $c = 0.207$ the optimum is $\alpha = 2.1682$ and the stability limit 2.1357. In any case we have

$$|g(\alpha, 0, 0, 0, 1)| > 0.033 > \left|g(2.13, 0.207, 1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})\right| \simeq 0.0078$$

Next let us plot the phase speed as a function of $2/N$ as before. First in figure 6.5 with $c = \varphi = \theta = 0$ (direction of the maximum error) we see that the error does not depend much on α as expected from (6.16). It is seen that some kind of optimal choice would be $\alpha \simeq 1.5$. Then in figure 6.6 the parameters are $c = 0.207$ and $\varphi = \theta = 45$ (again direction of the maximum error). When $c = 0$ the error increases with increasing α while it decreases when $c = 0.207$. It is seen that the error grows very fast when $N < 5$, so one has to use more points per wavelenght than when solving the ordinary wave equation. In figure 6.7 ($\varphi = \theta = 45$) we have then compared the different values of c, using $\alpha = 1.5$ for $c = 0$.

Then in figures 6.8 and 6.9 we have compared the accuracy of the scheme with $c = 0$ and $c = 0.207$ taking first $N = 5$ and then $N = 10$. Obviously $c = 0.207$ is the better value. However, taking $c \neq 0$ increases the amount of work. Now doing the same simple operation count as in (4.4) and scaling accordingly the work to be the same, we get the figures 6.10 and 6.11. The error with $N = 5$ is about the same in both cases, but when $N = 10$ the value $c = 0.207$ is again clearly better.

Next, passing to the direction of the group speed as usual we note that the scheme satisfies the conditions of proposition 1, so $\beta = O(h^4)$ as before. In figure 6.12 we have taken $N = 5$ and we see that the maximum error is about the same for $c = 0$ and $c = 0.207$. However, when $N = 10$ the error with $c = 0.207$ is considerably smaller, figure 6.13.

6.4 Polarization

Then how about the eigenvectors? We recall that for the continuous problem we had the orthogonality relations (6.7), which are numerically replaced by

$$\begin{aligned} \xi_c \cdot B_0 &= 0 \\ \xi_c \cdot D_0 &= 0 \end{aligned}$$

The plane spanned by B_0 and D_0 is called the plane of polarization and for the continuous problem it is orthogonal to k. Therefore a natural way to measure the error of polarization is to calculate

$$\gamma = \arccos\left(\frac{\xi_c \cdot k}{|\xi_c|\,|k|}\right)$$

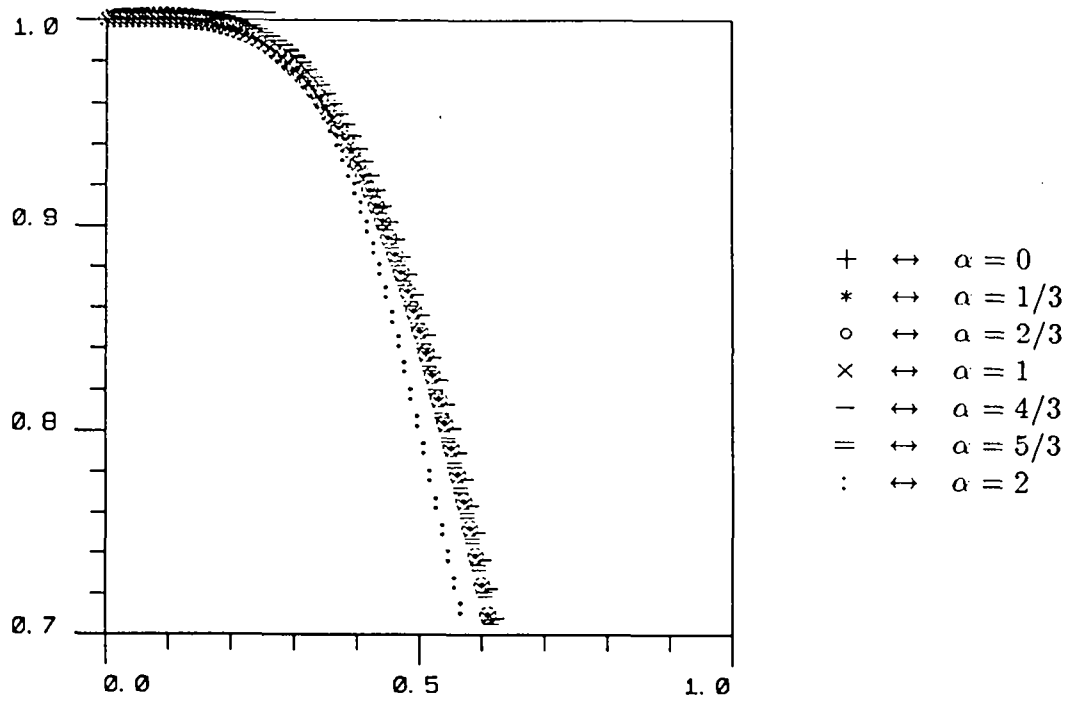


Figure 6.5: Phase speed: $c = \varphi = \theta = 0$.

Here we note that ξ_c satisfies the conditions (D1) - (D4), so with proposition 2 we conclude that $\gamma = O(h^4)$. This is true for any value of c, although ξ_c is only of second order if $c \neq 0$.

In figure 6.14 we show γ for some values of c and taking $\varphi = 40$ (other values of φ give similar results). The value $c = 0.207$ is clearly the best. In addition, comparing to 6.12 and 6.13 we see that typically $\gamma < \beta$.

Finally let us say a few words about the last two equations of (6.1). We remarked that they are consequences of the first two. Numerically the situation is perhaps less clear. However, consider the following discrete divergence operator.

$$\nabla_c \cdot \psi = \frac{1}{h}\left(S_x^c \psi_1 + S_y^c \psi_2 + S_z^c \psi_3\right)$$

Because the difference operators S^c commute we have

$$\nabla_c \cdot T_c \psi = 0$$

Then taking this discrete divergence of the equations (6.11) or (6.12) we see that numerical divergence is conserved.

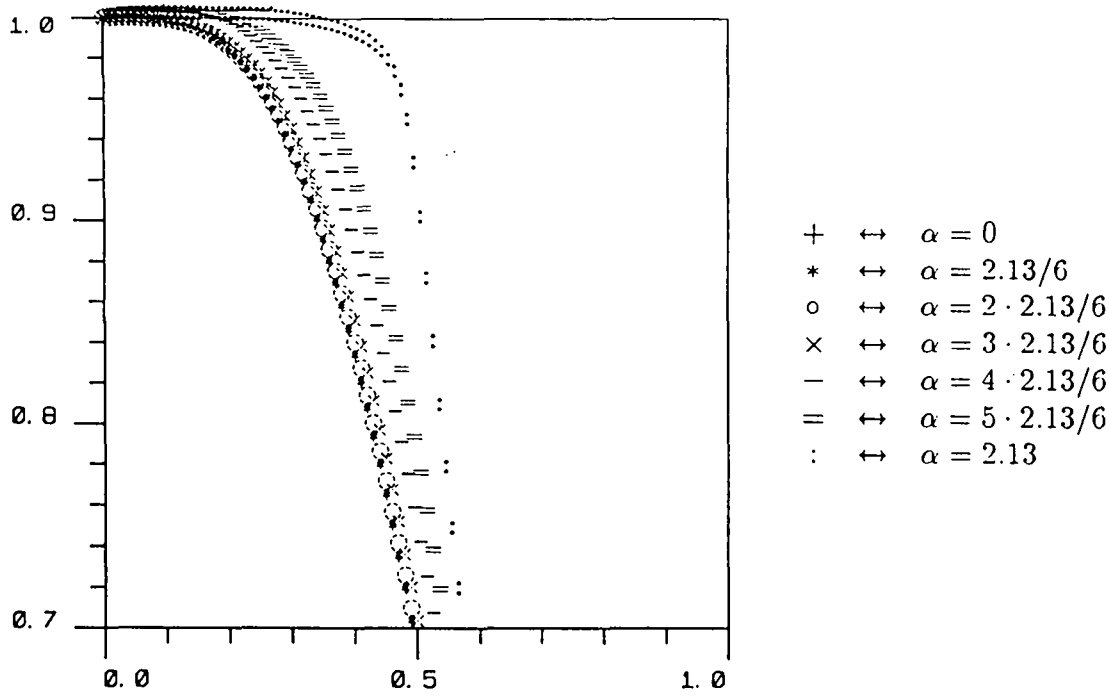


Figure 6.6: Phase speed: $\varphi = \theta = 45$, $c = 0.207$.

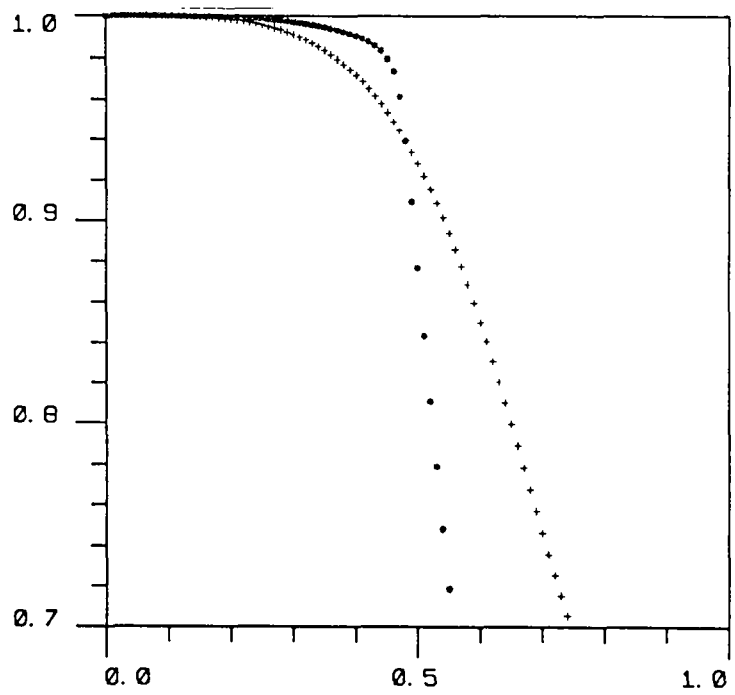


Figure 6.7: Phase speed: $+ \leftrightarrow c = 0$ and $\alpha = 1.5$; $* \leftrightarrow c = 0.207$ and $\alpha = 2.13$.

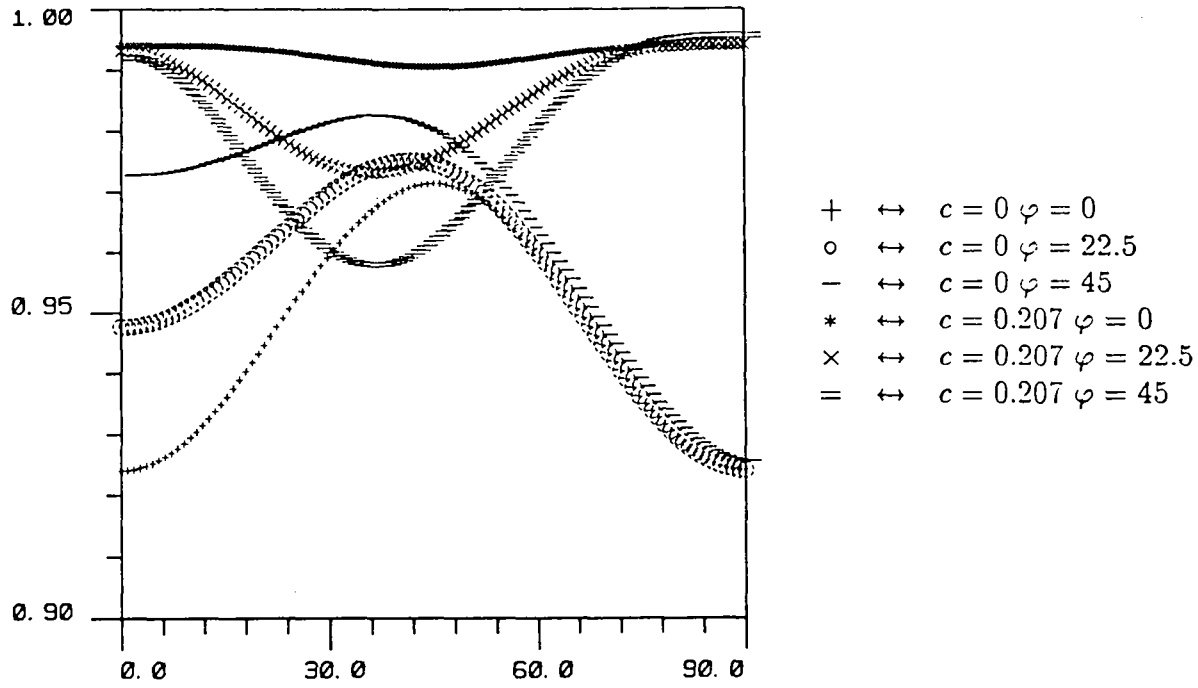


Figure 6.8: Phase speed without scaling, $N = 5$.

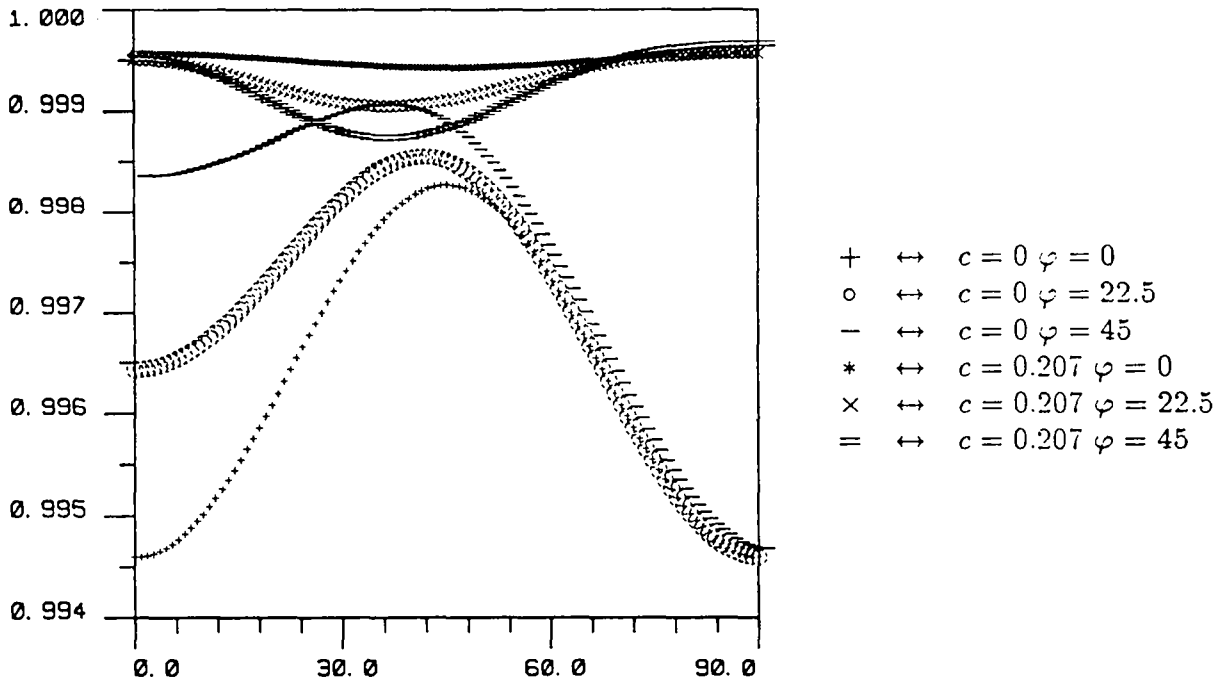


Figure 6.9: Phase speed without scaling, $N = 10$.

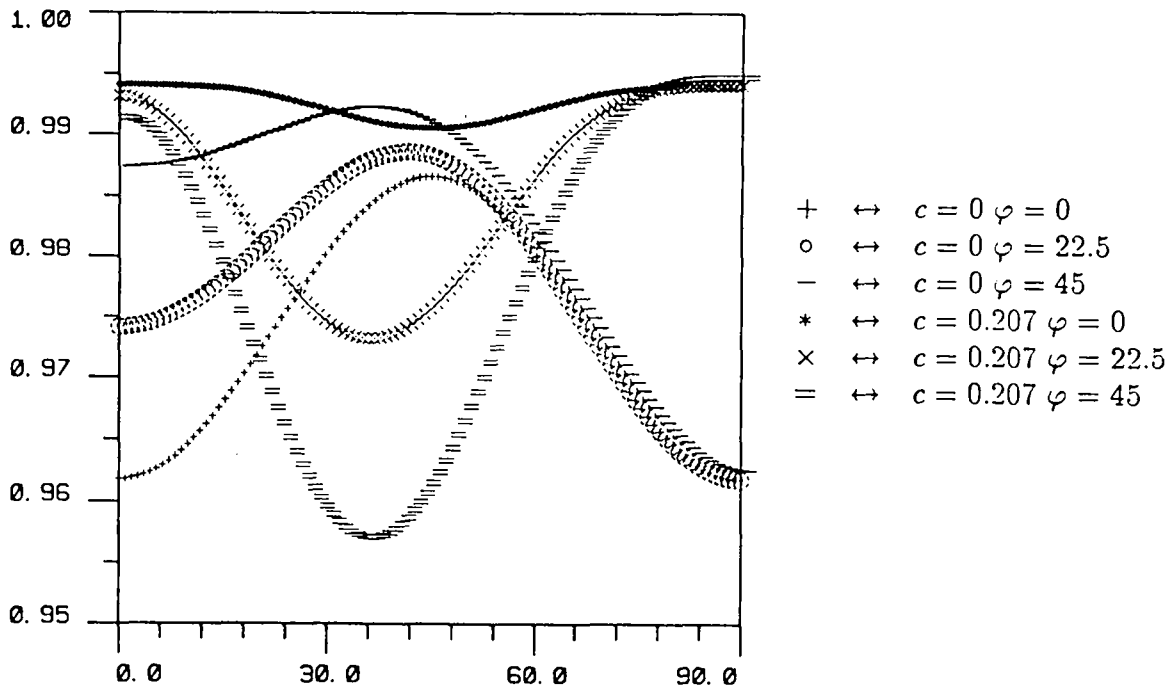


Figure 6.10: Phase speed with scaling, $N = 5$.

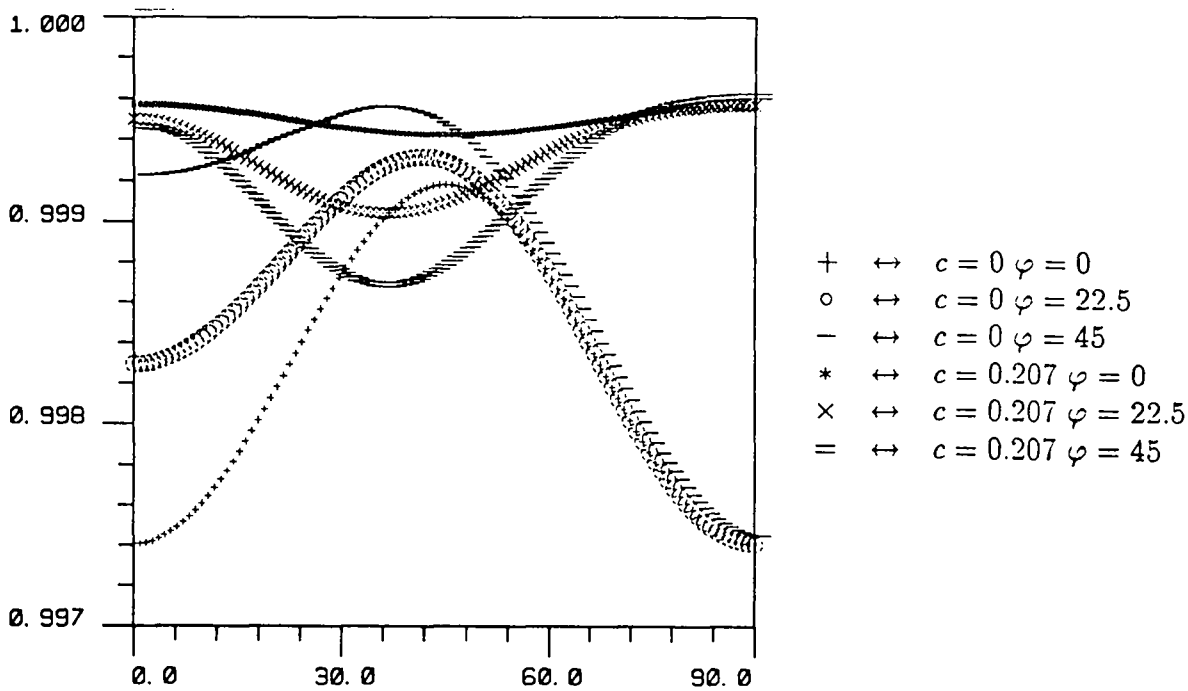


Figure 6.11: Phase speed with scaling, $N = 10$.

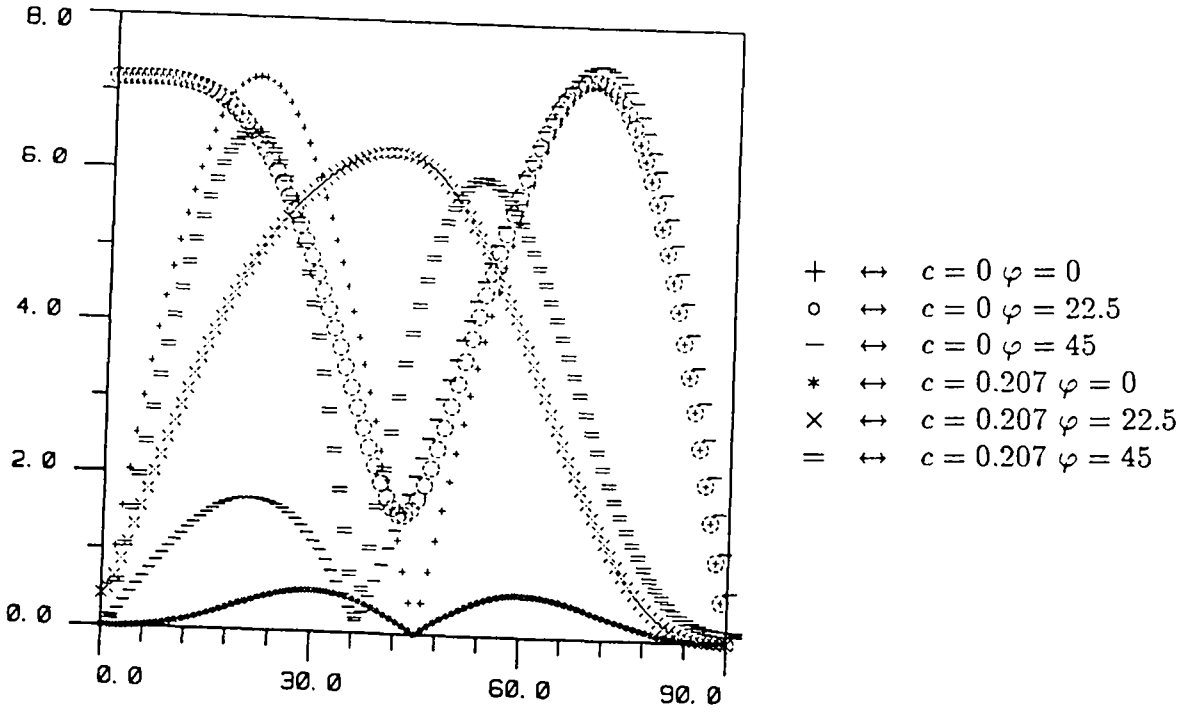


Figure 6.12: Error in the direction of propagation, $N = 5$

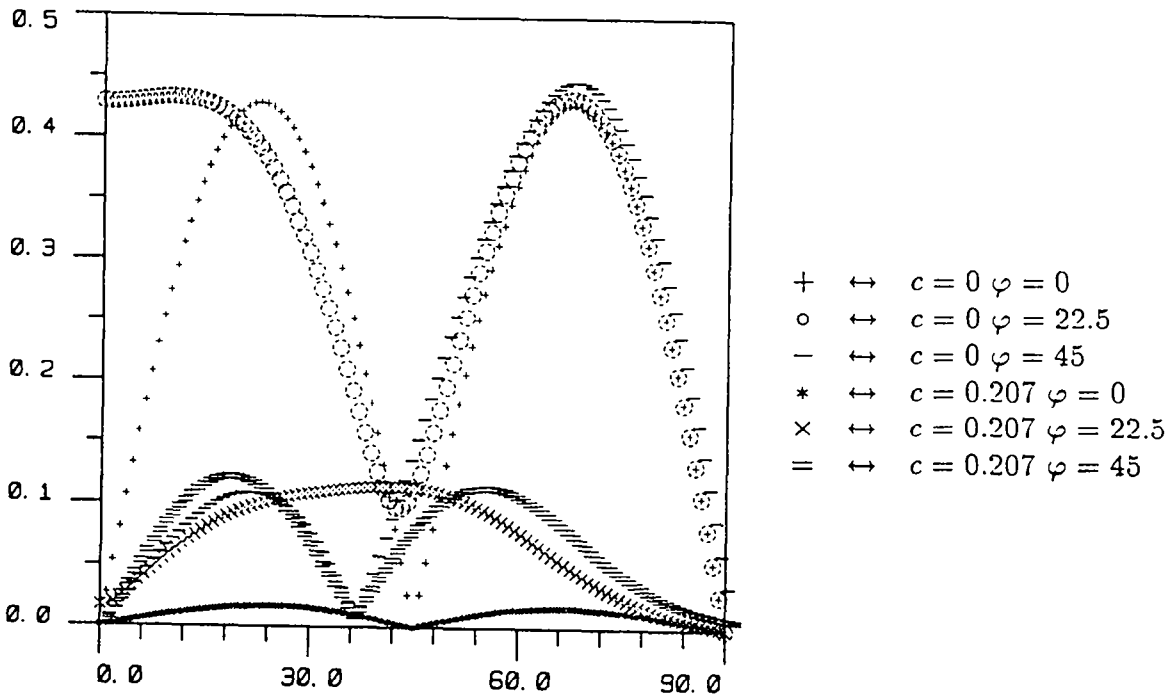


Figure 6.13: Error in the direction of propagation, $N = 10$

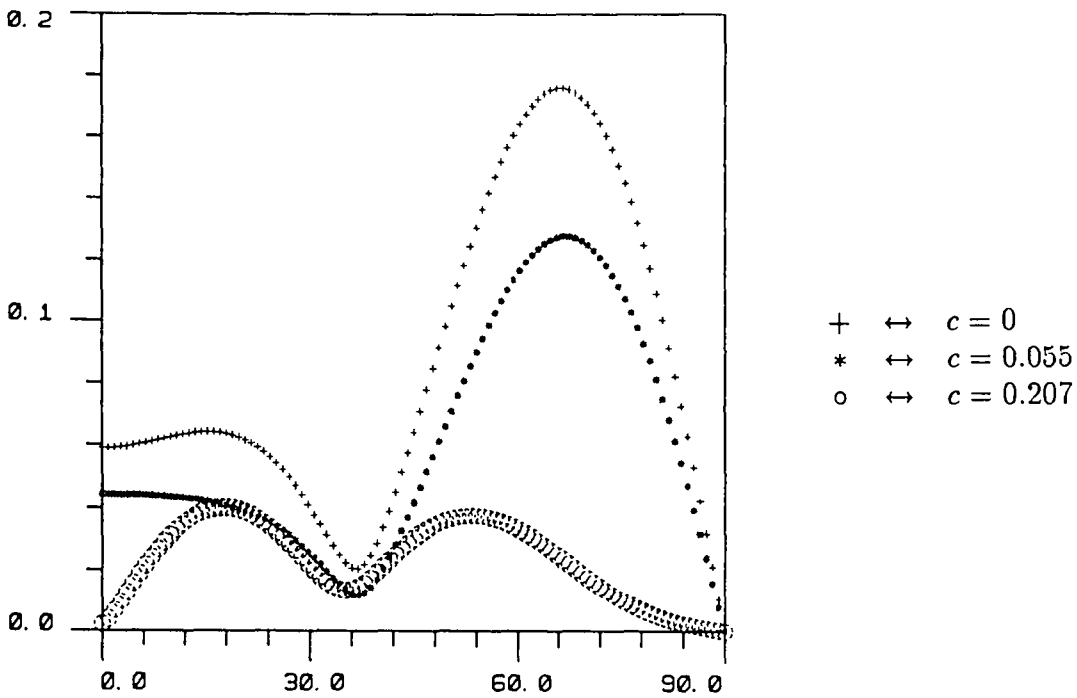


Figure 6.14: Polarization error, $N = 8$ and $\varphi = 40$.

7 Linearized Elastodynamic Equations

7.1 Background

The linearized elastodynamic equations can be written in various different ways, but for our purposes the most suitable form is the following.

$$u_{tt} - (\lambda + 2\mu)\Delta u - (\lambda + \mu)\nabla \times \nabla \times u = 0 \tag{7.1}$$

The parameters λ and μ are called the coefficients of Lamé and they are always positive (μ here has nothing to do with the permeability of (6.2)...). Searching for the plane waves of the form $u = u_0 \exp(i(k \cdot r - \omega t))$ as usual we get the following eigenvalue problem

$$\omega^2 u_0 = (\lambda + 2\mu)|k|^2 u_0 + (\lambda + \mu)K^2 u_0$$

The matrix K is here also given by (6.6). We have then two non zero solutions: a simple eigenvalue $\omega^2 = (\lambda + 2\mu)|k|^2$ and the double eigenvalue $\omega^2 = \mu|k|^2$. The eigenvector corresponding to the first eigenvalue is parallel to k, while the other two are orthogonal to it. The eigenvector parallel to k gives the solution called the pressure wave and its speed of propagation is $\sqrt{\lambda + 2\mu}$ and the other two solutions are the shear waves which travel with speed $\sqrt{\mu}$.

7.2 Construction of the Fourth Order Scheme

For the numerical solution of (7.1) we discretise the laplacian and $\nabla \times \nabla \times$ separately. For the latter we obviously use the operator T_c. For the laplacian we use the operator T_Δ defined in (5.1). So we obtain a second order scheme as follows.

$$\frac{u^{n+1} - 2u^n + u^{n-1}}{\delta t^2} + \frac{\lambda + 2\mu}{h^2} T_\Delta u^n - \frac{\lambda + \mu}{h^2} T_c^2 u^n = 0$$

Substituting a plane wave in the above equation gives the following eigenvalue problem

$$\frac{4\sin^2(\omega\delta t/2)}{\delta t^2} u_0 = \frac{(\lambda + 2\mu)}{h^2} f_\Delta u_0 + \frac{(\lambda + \mu)}{h^2} \mathcal{K}_c^2 u_0$$

We change the meaning of α slightly and define it to be $\delta t\sqrt{(\lambda + 2\mu)}/h$. Denote $(\lambda + \mu)/(\lambda + 2\mu)$ by b_1. Then the eigenvalue problem can be written as

$$4\sin^2(\omega\delta t/2)u_0 = \alpha^2 f_\Delta u_0 + b_1 \alpha^2 \mathcal{K}_c^2 u_0$$

Now when we want to construct a fourth order scheme we have exactly the same difficulty as before: we cannot directly use the fourth order discretisation of the time

derivative. So we write down the truncation error of the time discretisation in the usual way.

$$\begin{aligned}\frac{u^{n+1}-2u^n+u^{n-1}}{\delta t^2} &= \partial_t^2 u + \frac{\delta t^2}{12}\partial_t^4 u + O(\delta t^4) = \partial_t^2 u + \\ &\frac{\delta t^2}{12}((\lambda+2\mu)^2\Delta^2 u - (\lambda+\mu)(\lambda+3\mu)(\nabla\times)^4 u) + O(\delta t^4)\end{aligned}$$

Here we have used the identity $(\nabla\times)^4 = -\Delta(\nabla\times)^2$. The Taylor's expansion shows that

$$\begin{aligned}\frac{1}{h^2}T_\Delta u &= -\Delta u - \frac{h^2}{12}\Delta^2 u + O(h^4) \\ \frac{1}{h^2}T_c^2 u &= \nabla\times\nabla\times u - 2ch^2(\nabla\times)^4 u + O(h^4)\end{aligned}$$

So the fourth order scheme can be written in the following way.

$$\begin{aligned}&\frac{u^{n+1}-2u^n+u^{n-1}}{\delta t^2} + \frac{\lambda+2\mu}{h^2}T_\Delta u^n - \frac{\lambda+\mu}{h^2}T_c^2 u^n + \\ &\frac{(\lambda+2\mu)}{12h^4}(h^2-(\lambda+2\mu)\delta t^2)T_\Delta^2 u^n - \frac{(\lambda+\mu)}{12h^4}(24ch^2-(\lambda+3\mu)\delta t^2)T_c^4 u^n = 0\end{aligned}$$

Then using α, b_1 and $b_2 = (\lambda+3\mu)/(\lambda+2\mu)$, the above scheme can more conveniently be written as

$$\begin{aligned}&u^{n+1} - 2u^n + u^{n-1} + \alpha^2\big(T_\Delta u^n - b_1 T_c^2 u^n + \\ &\frac{1-\alpha^2}{12}T_\Delta^2 u^n - \frac{b_1}{12}(24c - b_2\alpha^2)T_c^4 u^n\big) = 0\end{aligned}$$

Substituting the plane wave as usual, the corresponding eigenvalue problem is then found to be

$$4\sin^2(\omega\delta t/2)u_0 = \alpha^2\left((f_\Delta + \frac{1}{12}(1-\alpha^2)f_\Delta^2)I + b_1(\mathcal{K}_c^2 - \frac{1}{12}(24c - b_2\alpha^2)\mathcal{K}_c^4)\right)u_0$$

Now there are two eigenvalues to approximate. However, the eigenvalue corresponding to the pressure wave depends only on the operator T_Δ, so it behaves exactly like the ordinary wave equation. This has already been analysed above so it is sufficient here to consider only the shear waves. In addition it is the shear waves that are difficult to approximate. This is because the wavelength of the pressure wave is at least $\sqrt{2}$ times bigger than the wavelength of the shear wave and consequently if there are sufficiently points to approximate well the shear wave the approximation of the pressure wave is even better. Note that also the above discrete problem has an

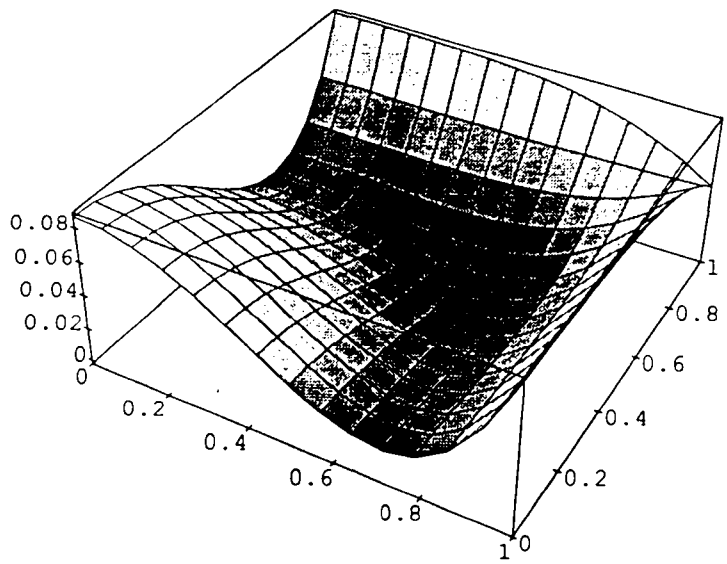


Figure 7.1: $\check{g}$ as function of a_1 and a_2

exact double eigenvalue, which is not the case with all schemes: one could have two distinct eigenvalues which converge to the same limit as h tends to zero.
For the shear waves we get the following dispersion relation.

$$4\sin^2(\omega\delta t/2) = \alpha^2\left(f_\Delta + \frac{1}{12}(1-\alpha^2)f_\Delta^2 - b_1(f_c + \frac{1}{12}(24c - b_2\alpha^2)f_c^2)\right) \qquad (7.2)$$

We have then the stability condition

$$0 \leq \alpha^2\left(f_\Delta + \frac{1}{12}(1-\alpha^2)f_\Delta^2 - b_1(f_c + \frac{1}{12}(24c - b_2\alpha^2)f_c^2)\right) \leq 4 \qquad (7.3)$$

It is clear that α cannot be bigger than 0.80, because $f_c = 0$ at the point where f_Δ attains its maximum value.

7.3 Optimising the Parameter c

Now c being a free parameter, how should we choose it? Let us make a Taylor's expansion of the relation (7.2) which can be written as

$$\omega = \sqrt{\mu}|k|(1 + \check{g}(\alpha, c, \lambda, \mu, a_1, a_2, a_3)h^4|k|^4 + O(h^6 + \delta t^6))$$

The expression for $\check{g}$ is too long to be written down, but proceeding as with equation (6.15), we can plot it for some values of c, α, λ and μ (like figures 6.3 and 6.4). We note that $\check{g}$ is homogeneous of degree zero with respect to λ and μ so only

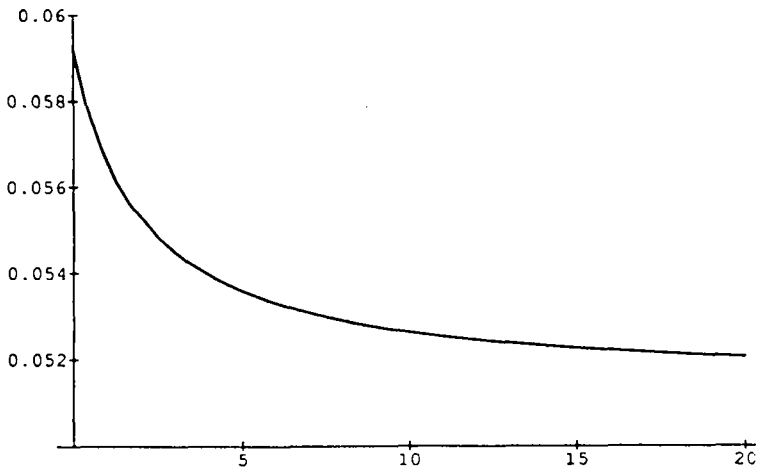


Figure 7.2: Optimal c as a function of λ/μ

the proportion λ/μ really matters. In figure 7.1 we show a typical situation with $\alpha = 0.80$ and $\lambda/\mu = 3$. We see that the maximum error is attained at origin, that is at $a_1 = a_2 = 0$. Now $\check{g}(\alpha, c, \lambda, \mu, 0, 0, 1)$ is a quadratic polynomial with respect to c, so one easily calculates that the minimum is at

$$c_{op} = \frac{(2\alpha^2 + 3)\lambda^2 + (8\alpha^2 + 9)\lambda\mu + (6\alpha^2 + 6)\mu^2}{84\lambda^2 + 252\lambda\mu + 168\mu^2}$$

In figure 7.2 we have plotted c_{op} as a function of λ/μ taking $\alpha = 0.80$ and in particular it is seen that $0.05 < c_{op} < 0.06$. In figure 7.3 there is $\check{g}(0.80, c, 3, 1, 0, 0, 1)$ as a function of c.

As regards the stability we first note that

$$24c_{op} - b_2\alpha^2 = \frac{6\lambda + 12\mu - (3\lambda + 9\mu)\alpha^2}{7\lambda + 14\mu} > 0.4$$

for all values of λ and μ and for $\alpha \leq 0.80$ so the coefficient of the f_c^2 term in (7.3) is positive and consequently we can conclude that the second inequality of (7.3) depends only on f_Δ. How about the first? This is more complicated but the numerical search confirms that it is also satisfied, when $c = c_{op}$ and $\alpha \leq 0.80$.

Next in figures 7.4, 7.5 and 7.6 we have plotted $\check{g}(\alpha, c_{op}, \lambda, \mu, 0, 0, 1)$ as a function of α for $\lambda/\mu = 0$, 3 and 20, respectively. We see that except for the small values of λ/μ the minimum is at $\alpha = 0.80$ and even then the difference is only about 0.001. More interesting is that the error grows rather dramatically when λ/μ grows. This can be seen also in pictures 7.7 and 7.8 (here also we use c_{op}), where the phase speed is

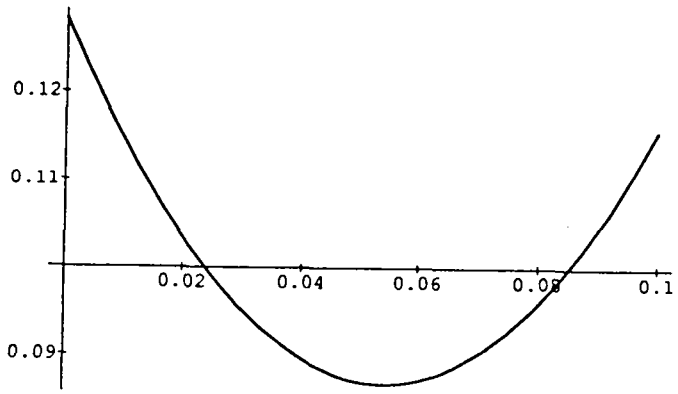


Figure 7.3: $\check{g}(0.80, c, 3, 1, 0, 0, 1)$

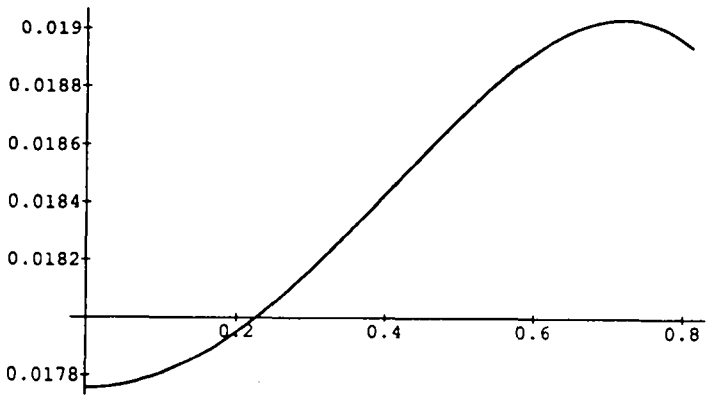


Figure 7.4: $\check{g}$ as a function of α, $\lambda/\mu = 0$.

plotted as a function of $2/N$ as usual. This can be interpreted as follows. If we keep λ fixed and let μ tend to zero, the speed of the numerical shear waves do not tend to zero (the right hand side of (7.2) will not become zero by putting $\mu = 0$). This means that the relative error grows without limit as $\mu \to 0$.

Next we present the phase speed curves with different values of N and it is seen that when $\lambda/\mu = 20$, then one needs at least $N = 15$ to obtain even satisfactory results.

Finally calculating the direction of the group speed is here more complicated than before because we have two different operators instead of one. In this case the direction is different for the compensated and uncompensated method. However, one easily verifies that in both cases the conditions of proposition 1 are satisfied so that the order remains the same. In the figures 7.15, 7.16, 7.17 and 7.18 there are then β

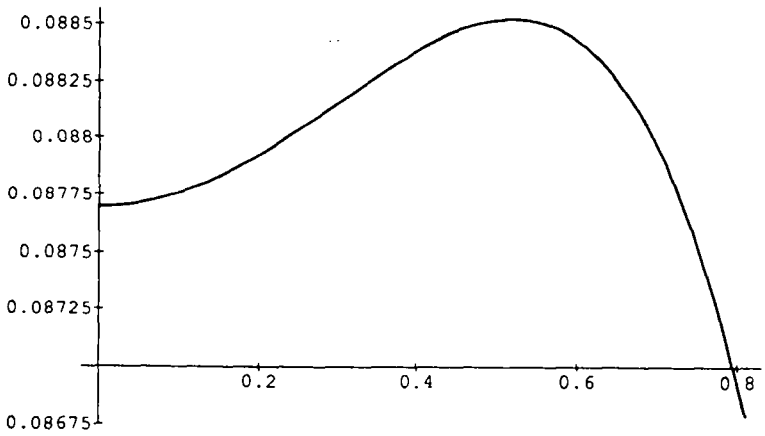


Figure 7.5: $\check{g}$ as a function of α, $\lambda/\mu = 3$.

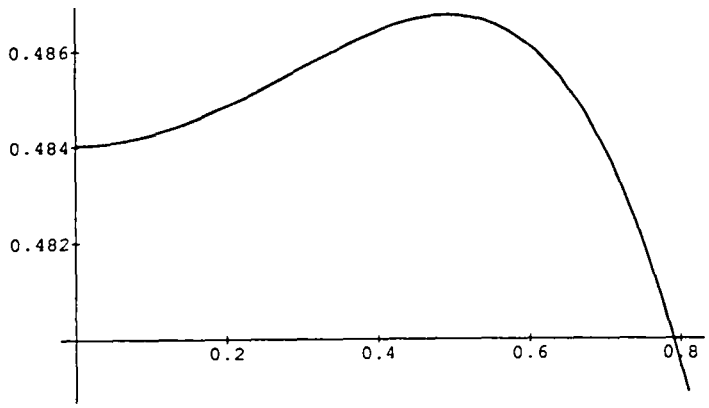


Figure 7.6: $\check{g}$ as a function of α, $\lambda/\mu = 20$.

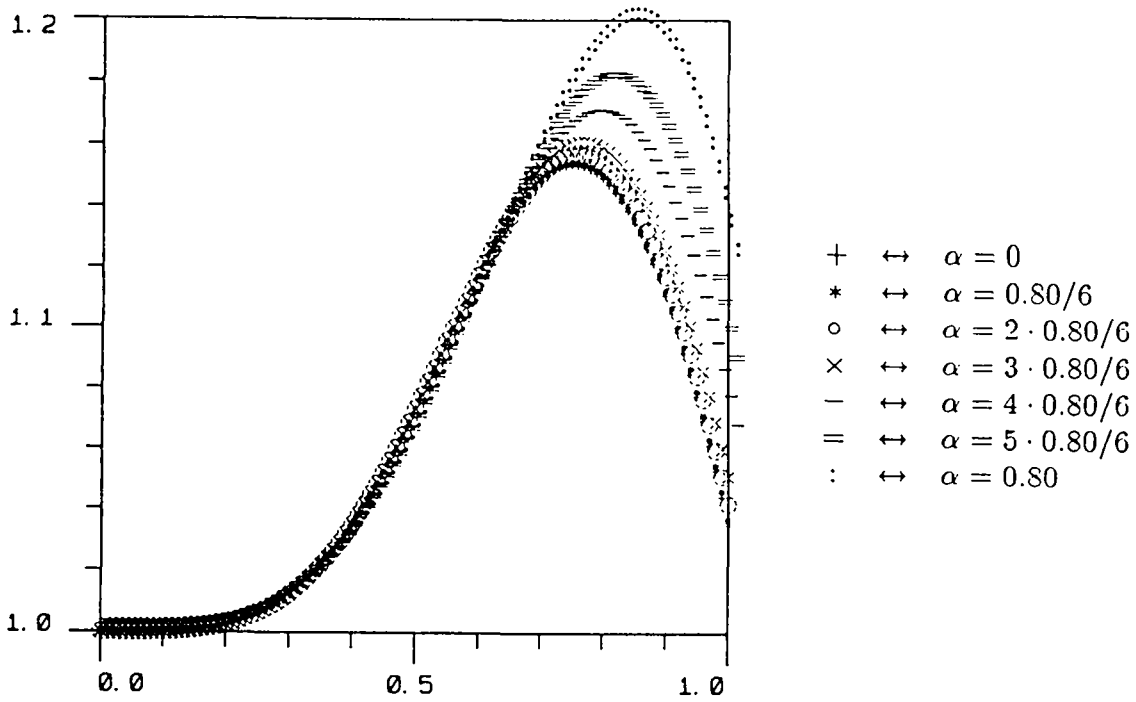


Figure 7.7: Phase speed, $\lambda/\mu = 0$ and $\varphi = \theta = 0$.

as a function of θ with different parameter values. The effect of λ/μ is as important as in the case of the phase speed.

As regards the polarization, it is seen that the eigenvectors for elastodynamic equations are exactly the same as for Maxwell's equations, because Δ and T_Δ are diagonal operators, so they do not change the direction of the eigenvectors.

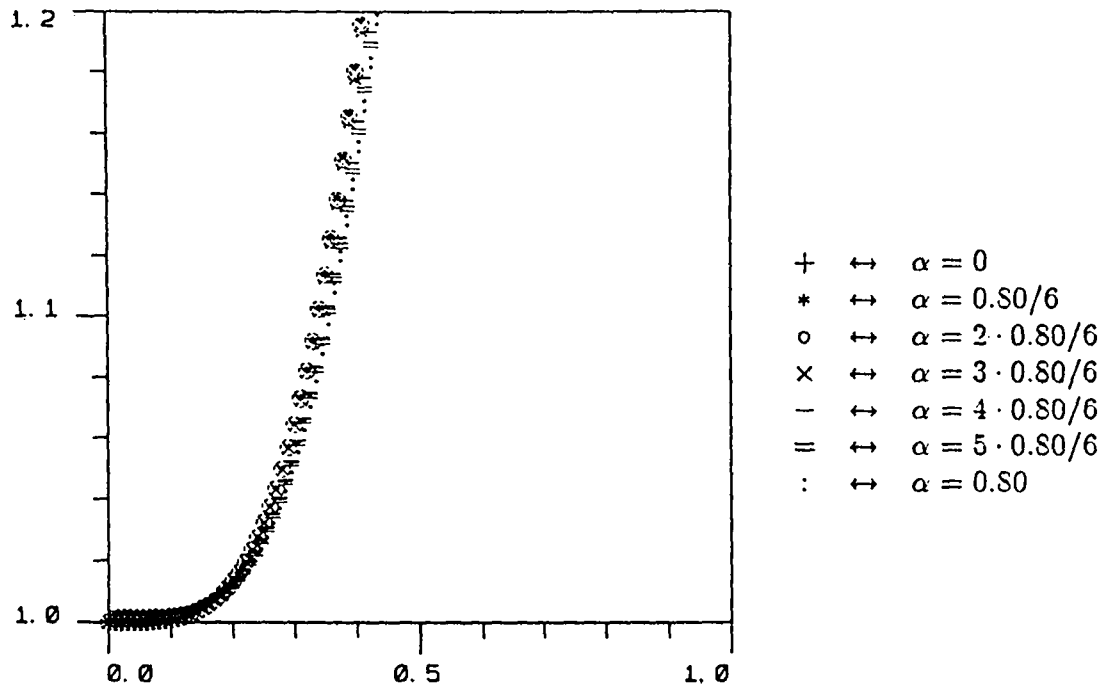


Figure 7.8: Phase speed, $\lambda/\mu = 3$ and $\varphi = \theta = 0$.

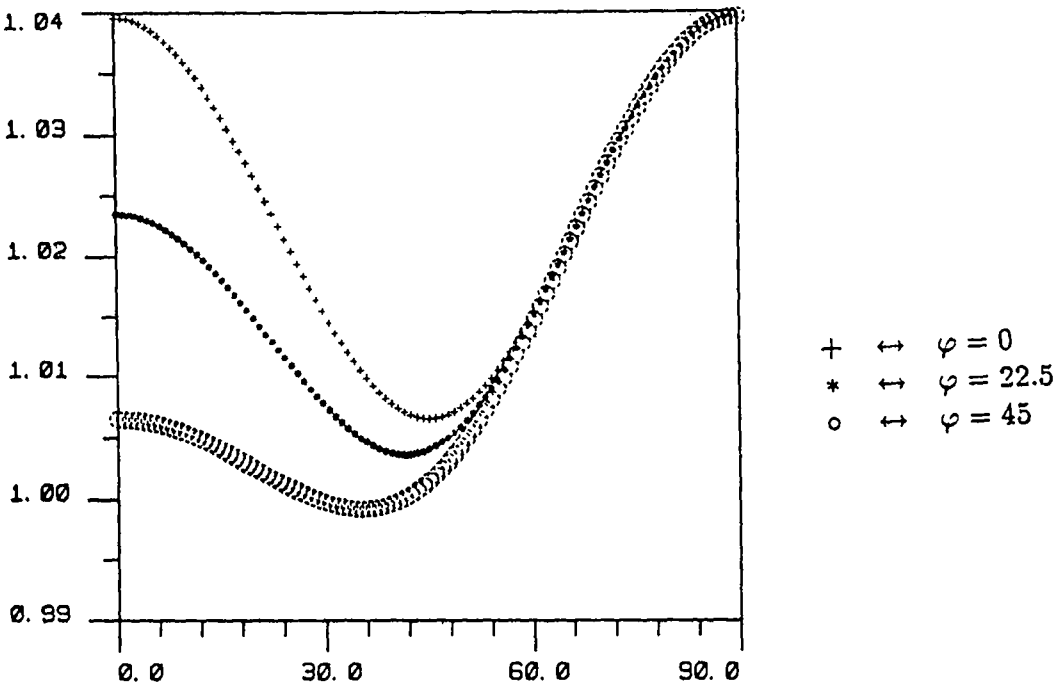


Figure 7.9: Phase speed, $\lambda/\mu = 0$, $N = 5$.

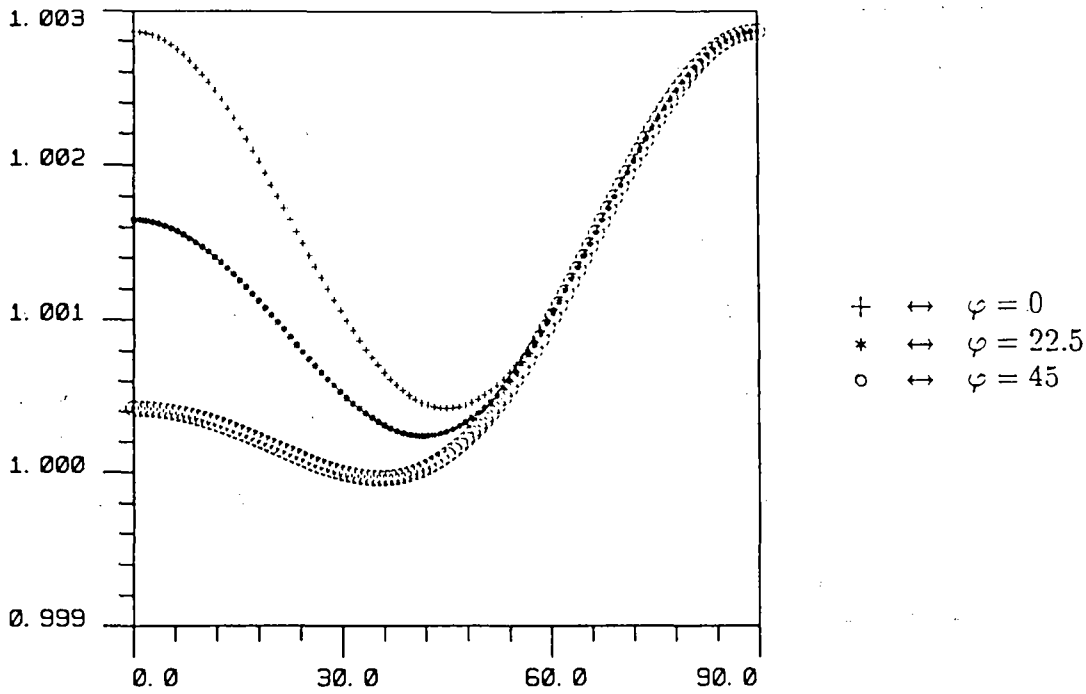


Figure 7.10: Phase speed, $\lambda/\mu = 0$, $N = 10$.

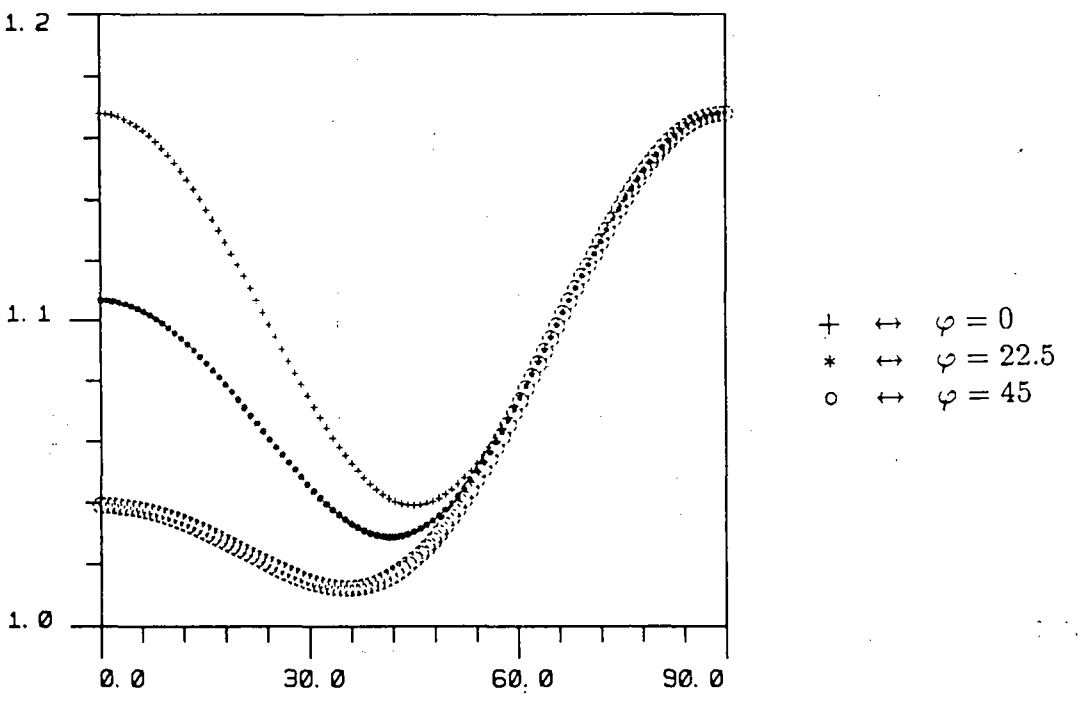


Figure 7.11: Phase speed, $\lambda/\mu = 3$, $N = 5$.

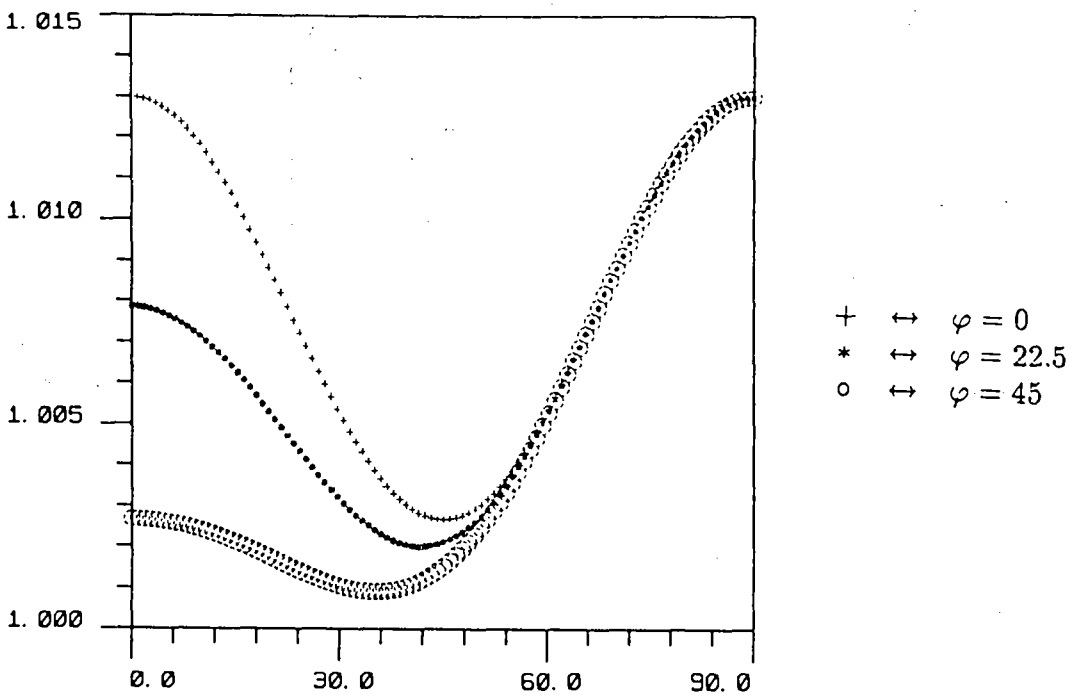


Figure 7.12: Phase speed, $\lambda/\mu = 3$, $N = 10$.

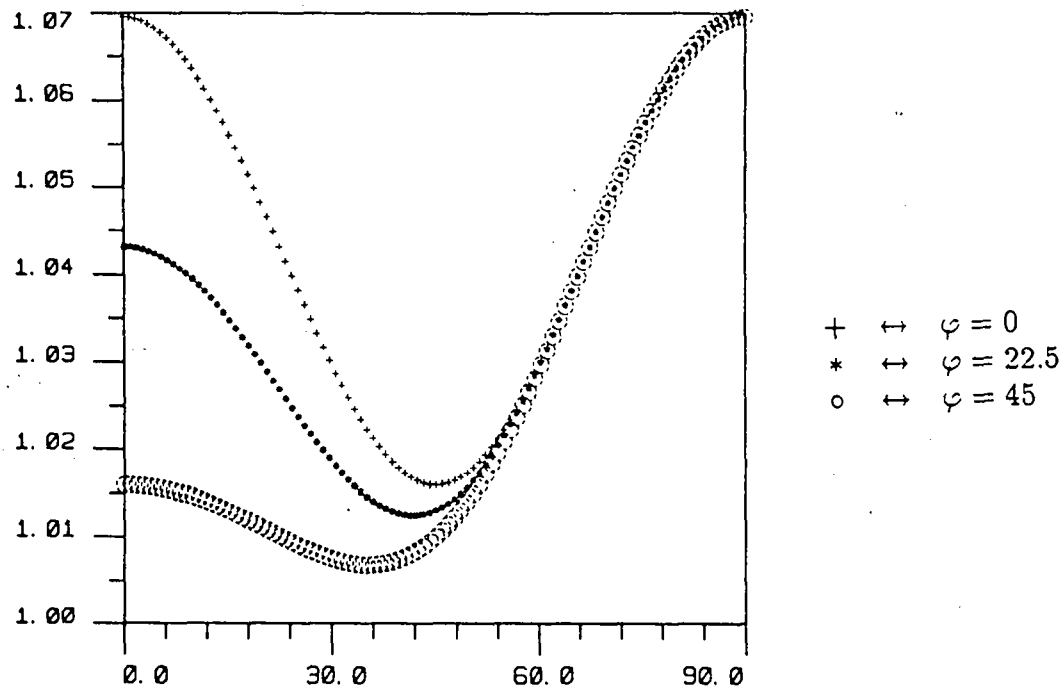


Figure 7.13: Phase speed, $\lambda/\mu = 20$, $N = 10$.

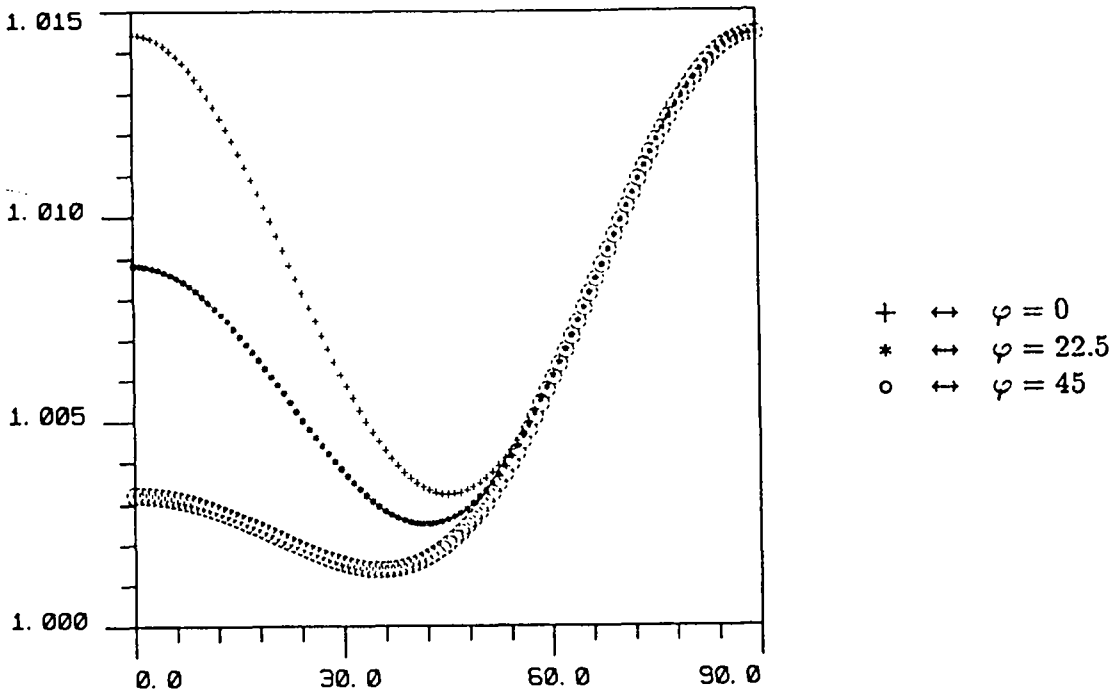


Figure 7.14: Phase speed, $\lambda/\mu = 20$, $N = 15$.

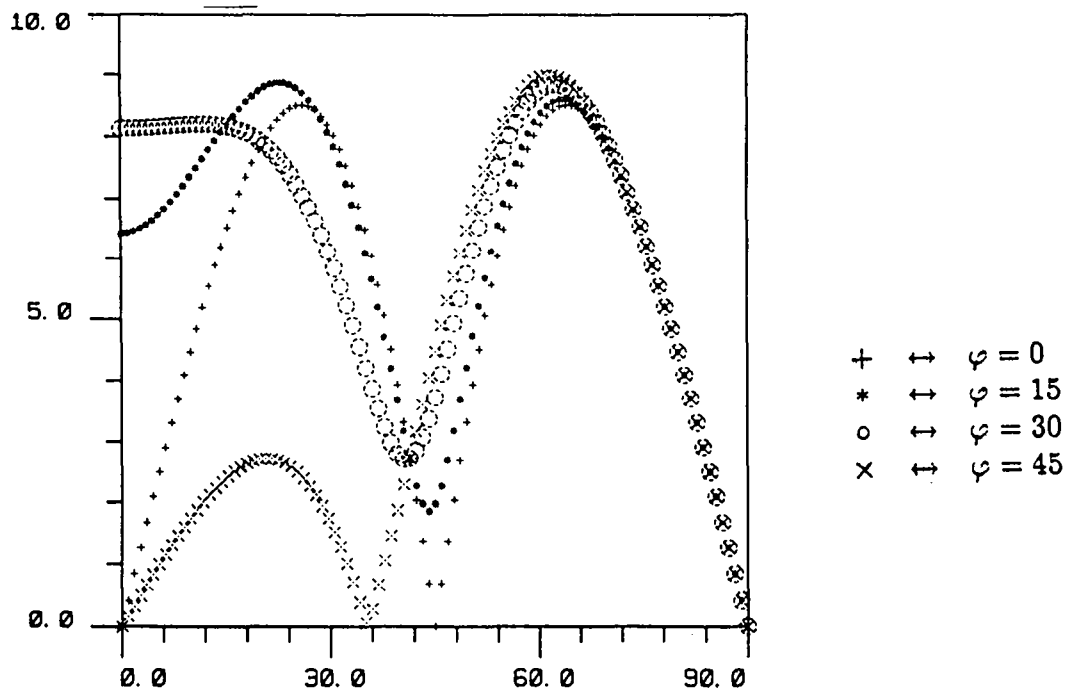


Figure 7.15: Error in the direction of propagation, $\lambda/\mu = 2$, $N = 5$.

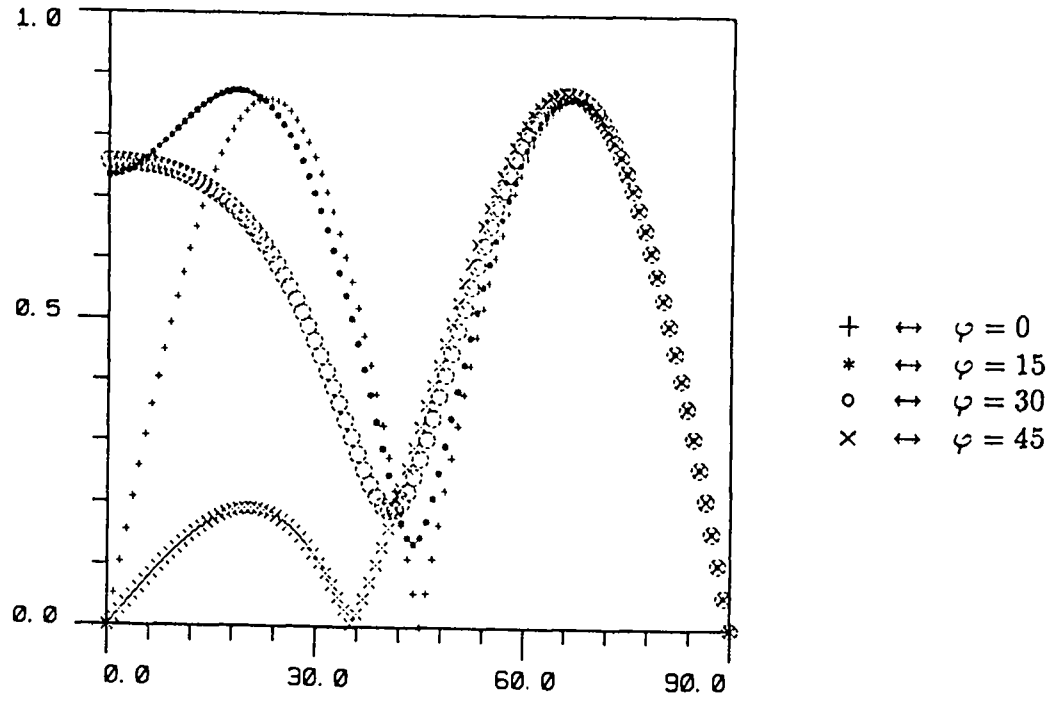


Figure 7.16: Error in the direction of propagation, $\lambda/\mu = 2$, $N = 10$.

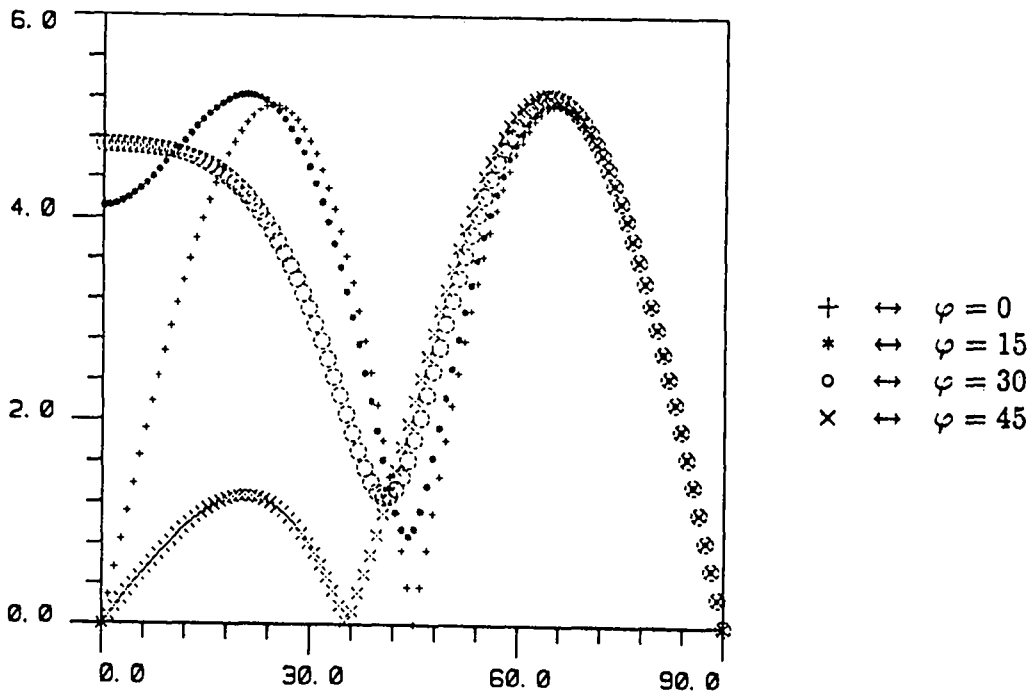


Figure 7.17: Error in the direction of propagation, $\lambda/\mu = 20$, $N = 10$.

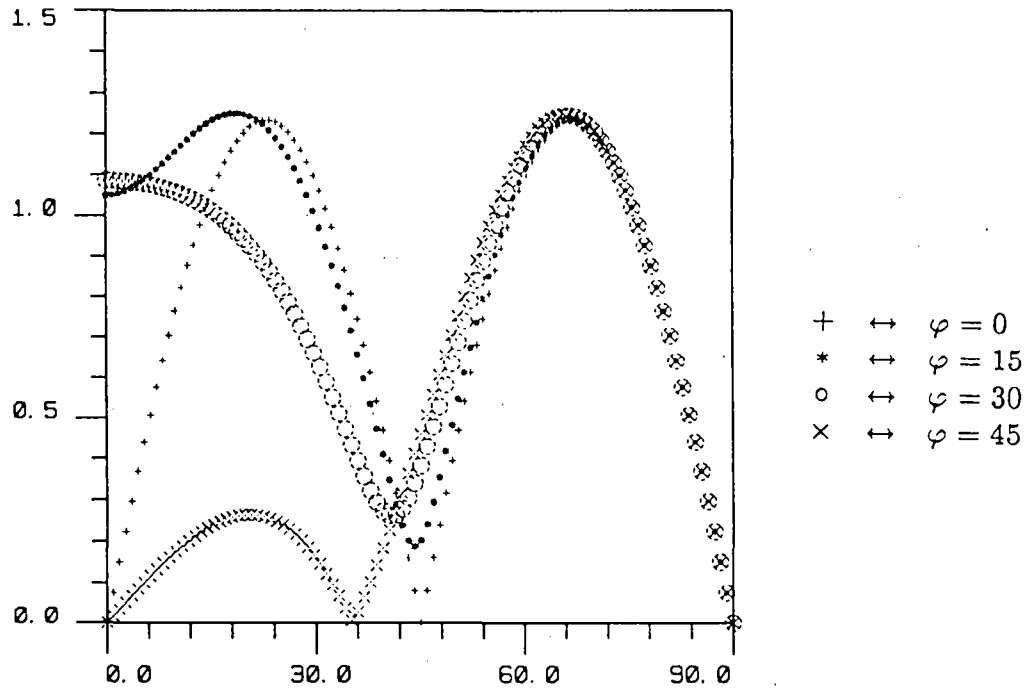


Figure 7.18: Error in the direction of propagation, $\lambda/\mu = 20$, $N = 15$.

8 Numerical Simulations

8.1 Test Problem

We have used the following test problem.

$$\begin{cases} u_{tt} - \Delta u = \delta(x) \otimes f(t) \\ u(x,0) = u_t(x,0) = 0 \end{cases} \tag{8.1}$$

where δ is Dirac's measure. We can calculate the analytical solution with the aid of the fundamental solution, if f is sufficiently smooth (for instance continuous). We recall that the fundamental solution for the wave operator is the solution of the following problem

$$u_{tt} - \Delta u = \delta(x) \otimes \delta(t)$$

and it is given by (for more details see [DL2])

$$E(|r|,t) = \begin{cases} \dfrac{1}{2\pi} \dfrac{1}{\sqrt{t^2 - |r|^2}} & \text{if } |r| < t \\ 0 & \text{if } |r| > t \end{cases}$$

Now the solution of (8.1) can be found using convolution.

$$u(|r|,t) = E * f = \int_0^t E(|r|,s) f(t-s) ds \tag{8.2}$$

Note that E is integrable when $r \neq 0$. It is possible to integrate (8.2) analytically in some cases but because the expressions usually are quite complicated and anyway they evidently are different for different f, we preferred to use the numerical integration (the Simpson's method). Supposing f to be coutinuously differentiable it is convenient to integrate (8.2) by parts (this facilitates the numerical integration) which gives for $t > |r|$

$$\begin{aligned} u(|r|,t) &= \frac{1}{2\pi} \int_{|r|}^t \ln(s - \sqrt{s^2 - |r|^2}) f'(t-s) ds \\ &+ \frac{1}{2\pi} \left(\ln|r| \, f(t-|r|) - \ln(t - \sqrt{t^2 - |r|^2}) f(0) \right) \end{aligned}$$

For $t < |r|$ u is of course zero.

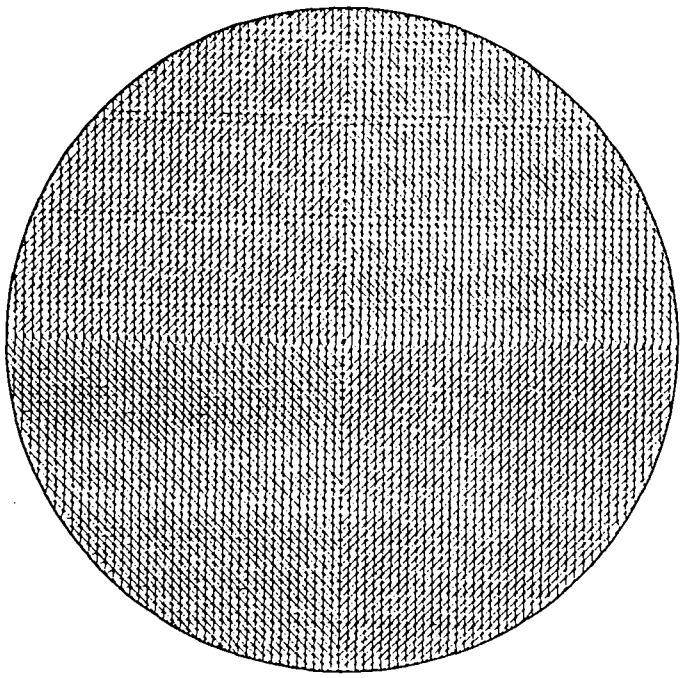


Figure 8.1: Domain for finite difference method

8.2 Results

Now when deriving the fourth order schemes we have not taken into account the right hand side. If this were done then instead of the modified equation (4.1), we would have the following 'modified equation with modified right hand side'.

$$u_{tt} - \Delta u - \frac{\delta t^2}{12}\Delta^2 u = f + \frac{\delta t^2}{12}(f_{tt} + \Delta f)$$

Shortly we will see that this omission is of some importance.

In figure 8.2 it is shown the domain and the triangulation for the P1 elements and in figure 8.1 the same for the finite difference method (2.4). In the following cases we consider the solutions only before they reach the boundary, so the boundary conditions do not have any effect on results (for the same reason it does not matter that the two domains are not exactly equal). However, all the usual boundary conditions (Dirichlet, Neumann, absorbing etc) could be imposed and in figure 8.3 we show a typical situation with absorbing boundary conditions.

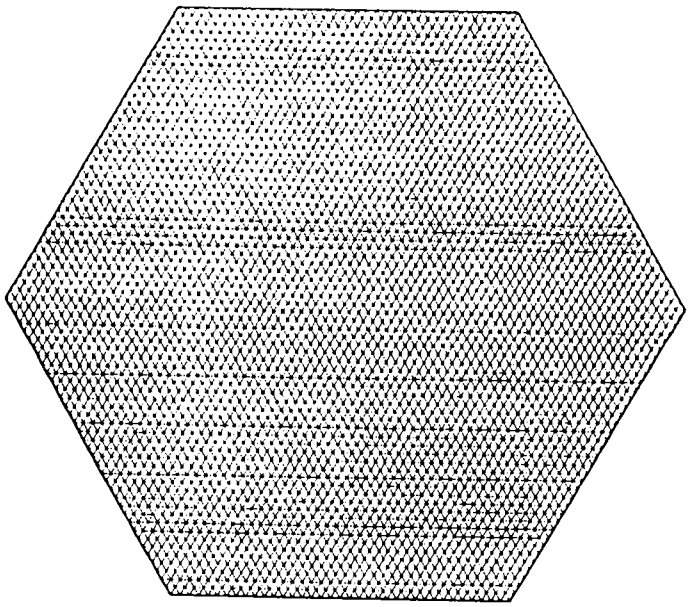


Figure 8.2: Domain for P1 elements

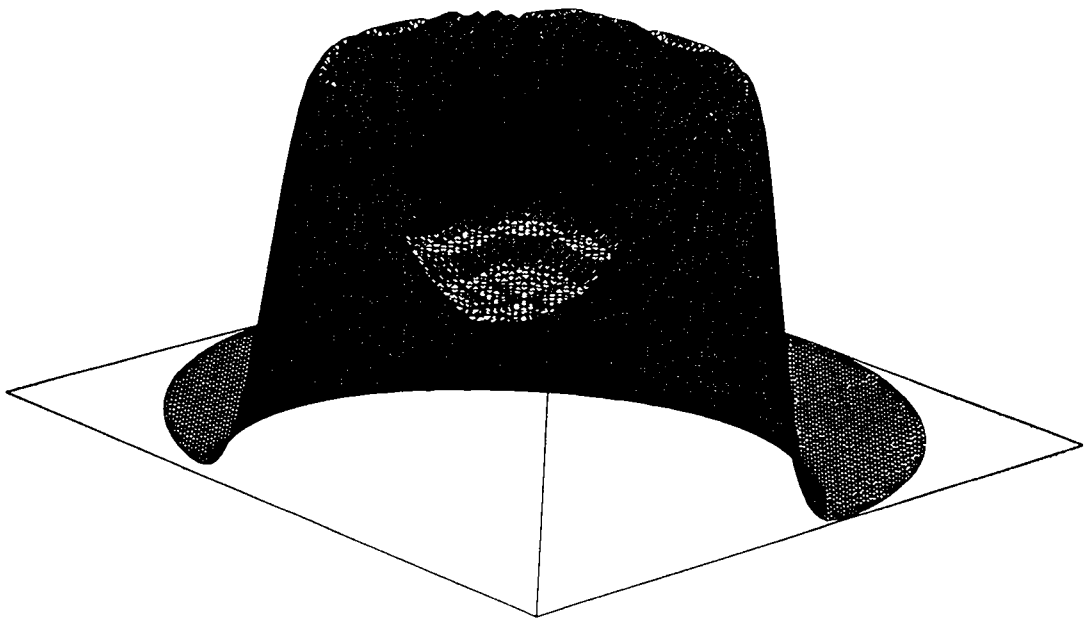


Figure 8.3: A typical solution

h	δt	t_0	error	N
0.28	0.2	0.0	$3.64 \cdot 10^{-1}$	7
0.14	0.1	0.0	$6.40 \cdot 10^{-2}$	14
0.28	0.2	2.4	$1.18 \cdot 10^{-1}$	7
0.14	0.1	2.4	$7.34 \cdot 10^{-3}$	14

Table 1: Accuracy of the compensated method

Let us first consider the following right hand side.

$$f(t) = \begin{cases} \sin^4(\pi t/b) & 0 \le t \le b \\ 0 & \text{otherwise} \end{cases}$$

Note that $f \in C^3(\mathbb{R})$ which implies that $u \in C^4(\mathbb{R}^2 \times \mathbb{R})$ when $t > b$. The number b is here a kind of nominal wavelength or pulse width and we recall that f 'contains' infinitely many and arbitrarily large frequencies (Fourier transform of a compactly supported function cannot have a compact support). Thus the parameter N below is interpreted as the number of points per pulse width: $N = b/h$.

Then using the method (4.2) we show the results for this f in the table 1, first taking the initial time t_0 to be $t_0 = 0$ and then $t_0 = 2.4$. The latter avoids the problem of 'modified right hand side' mentioned above. The final time $\bar{t}$ was taken in both cases to be $\bar{t} = 8$. The discrete L_∞ norm was used to calculate the error. At $\bar{t} = 8$, the L_∞ norm of the exact solution is about 12, so the relative error is about one order of magnitude smaller than the absolute error. It is seen that when $t_0 = 2.4$ the scheme is 'truly' a fourth order method since $0.118/7.34 \cdot 10^{-3} \simeq 16.1$, while when $t_0 = 0$ we get only $0.364/0.064 \simeq 5.7$. So it could be a good idea to modify also the right hand side, at least when it is smooth. However, in practise the source terms might be known only pointwise or be otherwise irregular which could make the implementation difficult. Next we will compare the compensated method to the ordinary finite difference scheme (2.4) (which is equivalent to P1 elements with rectangular grid and mass lumping), taking f to be less regular. We chose an

scheme	h	δt	error	CPU	N
fd2	0.25	0.17	0.63	3.8	8
P1 2	0.28	0.2	0.45	2.9	7
P1 4	0.28	0.2	0.17	4.7	7
fd2	0.125	0.085	0.18	34.5	16

Table 2: Comparison of second order and fourth order methods.

ordinary cubic spline, that is

$$f(t) = \begin{cases} 3t^2/b^2 - 2t^3/b^3 & 0 \le t \le b \\ 1 - 3(t-b)^2/b^2 + 2(t-b)^3/b^3 & b \le t \le 2b \\ 0 & \text{otherwise} \end{cases}$$

The pulse width is now $2b$ and consequently $N = 2b/h$. The results can be seen in table 2 (with $t_0 = 0$ and $\bar{t} = 8$). The norm of the error is the same as before; here also the relative error were about one tenth of the absolute error. CPU gives the time in seconds needed to compute the solution without the overhead. The computations were done on Apollo dn 10000 work station. It is seen that to obtain the same accuracy with the scheme (2.4) (denoted by fd2) one has to use $34.5/4.7 \simeq 7.3$ times more CPU time than with the scheme (4.2) (denoted by P1 4, P1 2 is the scheme (2.6)).

The results indicate that it is preferable to use fourth order schemes even when the solution is not smooth enough. In other words fourth order schemes are superior to second order schemes even when the error does not behave like $O(h^4)$. In the example above the error was seen to be $O(h^2)$ also for the fourth order scheme. This can be interpreted in the following way. Normally the singularities of the solution are concentrated in small parts of the domain. These singularities then reduce the global convergence rate, but since the solution is smooth in most parts of the domain it still helps to use a fourth order scheme.

9 Conclusion

Finally we compare our approach to some other methods which have been proposed, namely [HO] and [CJ]. Holberg constructs some special difference operators for approximating the second derivative (to solve the wave equation). These operators are long, that is they require quite many function evaluations. On the other hand, because they are long, there are a lot of free parameters, which can then be chosen so that the numerical dispersion is very small. Of course this applies only to the space discretisation because of the stability problem discussed above and so he must use the usual three point scheme in time. This means that the time step has to be very small, typically $\delta t = O(h^2)$, otherwise the error in time discretisation destroys the good properties of the space discretisation.

Cohen and Joly in [CJ] construct schemes which are of the fourth order even for the non homogeneous domains. However, the variation of the local wave speed has to be smooth. Their methods being more general than (but not including) the ones developed above they typically would require more computer time and/or storage space. The accuracy should be similar in situations were both approaches are applicable, but we have not made any comparisons.

Acknowledgements. Many of the calculations above would have been very tedious or (almost) impossible if done by hand. Therefore we have used all along the program Mathematica for doing symbolic computation (it does also numerical computations and graphics). Mathematica was developed by Wolfram and his collaborators and it is described in [WO]. In particular, the phrases like 'it is straightforward to...' used in this report mean that you can easily do it with Mathematica (if you know how to use it), but otherwise it might take quite a lot of time. I would like to thank professor Branko Grünbaum for pointing me out the references [GO] and [GS] and Patrick Joly for many interesting discussions on dispersion and wave propagation.

References

[BCL] A. Bamberger, G. Chavent, P. Lailly: *Etude de schémas numériques pour les équations de l'élastodynamique linéaire*; rapport INRIA, 41, 1980.

[CJ] G. Cohen, P. Joly: *Schémas d'ordre quatre en temps et en espace pour l'équation des ondes acoustiques*; rapport INRIA, to appear.

[CO] G. Cohen: *Fourth order schemes for elastic wave propagation*; rapport INRIA, to appear.

[DL1] R. Dautray, J. L. Lions: *Analyse mathématique et calcul numérique, vol 6*; Masson, 1988.

[DL2] R. Dautray, J. L. Lions: *Analyse mathématique et calcul numérique, vol 7*; Masson, 1988.

[GO] M. Goldberg: *Three infinite families of tetrahedral space-fillers*; J. Combinatorial Theory (A), 16, 348-354, 1974.

[GS] B. Grünbaum, G. C. Shephard: *Tilings with congruent tiles*; Bulletin of AMS, vol 3, n 3, 1980.

[HJ] R. Horn, C. Johnson: *Matrix analysis*; Cambridge university press, 1985.

[HO] O. Holberg: *Computational aspects of the choice of operator and sampling interval for numerical differentiation in large-scale simulation of wave phenomena*; Geophysical Prospecting, 35, 629-655, 1987.

[JO] P. Joly: *Analyse numérique et mathématique de problèmes liés à la propagation d'ondes acoustiques, élastiques et électromagnetiques*; thèse de l'univérsite de Paris 9, 1987.

[KO] J. A. Kong: *Electromagnetic waves*; Wiley, 1986.

[LL] L. Landau, E. Lifschitz: *La théorie des champs*; Mir, 1970.

[RT] P. A. Raviart, J. M. Thomas: *Introduction à l'analyse numérique des équations aux dérivées partielles*; Masson, 1983.

[SB] G. R. Shubin, J. B. Bell: *A modified equation approach to constructing fourth order methods for acoustic wave propagation*; SIAM J. Sci. Stat. Comp. 135-151, vol 8, n 2, 1987.

[TR] L. N. Trefethen: *Group velocity in finite difference schemes*; SIAM Review, 24, 1982.

[WO] S. Wolfram: *Mathematica: a system for doing mathematics by computer*; Addison-Wesley, 1988.

[ZI] O. Zienkiewicz: *The finite element method*; 3rd edition, McGraw-Hill, 1977.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique

ISSN 0249-6399